

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number  
**WO 02/079395 A2**

- (51) International Patent Classification<sup>7</sup>: **C12N** 20 Gideons Point Road, Tonka Bay, MN 55331 (US).  
**SCHROEDER, William, A.** [US/US]; 3509 Highlands Road, Brooklyn Park, MN 55443 (US).
- (21) International Application Number: PCT/US02/02124
- (22) International Filing Date: 25 January 2002 (25.01.2002) (74) Agent: **DEGRANDIS, Paula**; Cargill, Incorporated, P.O. Box 5624, Minneapolis, MN 55440-5624 (US).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/264,329 26 January 2001 (26.01.2001) US  
60/288,984 4 May 2001 (04.05.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:  
US 60/264,329 (CIP)  
Filed on 26 January 2001 (26.01.2001)  
US 60/288,984 (CIP)  
Filed on 4 May 2001 (04.05.2001)
- (71) Applicant (for all designated States except US):  
**CARGILL, INCORPORATED** [US/US]; P.O. Box 5624, Minneapolis, MN 55440-5624 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **DE SOUZA, Mervyn, L.** [US/US]; Apartment 10, 1029 Raymond Avenue, St. Paul, MN 55114-1134 (US). **KOLLMANN, Sherry, R.** [US/US]; 12031 99th Avenue North, Maple Grove, MN 55369 (US). **MAY, Colleen, A.** [US/US];
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



**WO 02/079395 A2**

(54) Title: CAROTENOID BIOSYNTHESIS

(57) Abstract: Membranous bacteria that produce astaxanthin and other carotenoids are described, as well as isolated nucleic acids and expression vectors that can be used for producing carotenoids in microorganisms.

## Carotenoid Biosynthesis

### TECHNICAL FIELD

The invention relates to methods and materials for producing carotenoids, and in particular, to nucleic acid molecules, polypeptides, host cells, and methods that can be used for producing carotenoids.

### BACKGROUND

5 Astaxanthin (3,3'-dihydroxy- $\beta,\beta$ -carotene-4,4'-dione) is the primary carotenoid that imparts the pink pigment to the eggs, flesh, and skin of salmon, trout, and shrimp. Most animals cannot synthesize carotenoids. Rather, the pigments are acquired through the food chain from marine algae and phytoplankton, the primary producers of  
10 astaxanthin. ATX exists in three configurational isomers [(3S, 3'S), (3R, 3'R) and (3S, 3'R; 3R, 3'S)], however, ATX is found in the marine environment only in the (3S, 3'S) form. Consequently, this form is considered the natural and most desirable form of ATX.

Although astaxanthin has been commercially extracted from some yeast and crustacea species and has been chemically synthesized as a 1:2:1 mixture of the (3S,3'S)-,  
15 (3S,3'R)- and (3R,3'R)-isomers, astaxanthin is limited in availability and is expensive to purchase. See, Torrisen et al. (1989) Crit. Rev. Aquatic Sci. 1:209; and Mayer (1994) Pure Appl. Chem., 66:931-938. Thus, there is a need for a less expensive source of the naturally-occurring (3S,3'S) astaxanthin.

### SUMMARY

20 The invention is based on methods and materials for producing carotenoids such as lycopene, zeaxanthin, zeaxanthin diglucoside, canthaxanthin,  $\beta$ -carotene, lutein, and astaxanthin. Such carotenoids can be used as nutritional supplements in humans and can be formulated for use in aquaculture or as an animal feed. The invention provides nucleic acid molecules that can be used to engineer host cells having the ability to produce  
25 particular carotenoids and polypeptides that can be used in cell-free systems to make particular carotenoids. The engineered cells described herein can be used to produce large quantities of carotenoids.

In one aspect, the invention features an isolated nucleic acid having at least 76% sequence identity to the nucleotide sequence of SEQ ID NO:1 (e.g., at least 80%, 85%, 90%, or 95% sequence identity to the nucleotide sequence of SEQ ID NO:1) or to a fragment of SEQ ID NO:1 at least 33 contiguous nucleotides in length. An isolated  
5 nucleic acid can encode a zeaxanthin glucosyl transferase polypeptide at least 75% identical to the amino acid sequence of SEQ ID NO:2. Expression vectors containing such nucleic acids operably linked to an expression control element also are featured.

In another aspect, the invention features an isolated nucleic acid having at least 78% sequence identity to the nucleotide sequence of SEQ ID NO:3 (e.g., at least 80%,  
10 85%, 90%, or 95% sequence identity to the nucleotide sequence of SEQ ID NO:3) or to a fragment of SEQ ID NO:3 at least 32 contiguous nucleotides in length. An isolated nucleic acid can encode a lycopene  $\beta$ -cyclase polypeptide at least 83% identical to the amino acid sequence of SEQ ID NO:4.  $\beta$ -carotene can be made by contacting lycopene with a polypeptide encoded by such isolated nucleic acids. The invention also features an  
15 expression vector that includes such nucleic acids operably linked to an expression control element.

In yet another aspect, the invention features an isolated nucleic acid having at least 81% sequence identity to the nucleotide sequence of SEQ ID NO:5 (e.g., at least 85%, 90%, or 95% sequence identity to the nucleotide sequence of SEQ ID NO:5) or to a  
20 fragment of SEQ ID NO:5 at least 60 contiguous nucleotides in length. An isolated nucleic acid also can encode a geranylgeranyl pyrophosphate synthase polypeptide at least 85% identical to the amino acid sequence of SEQ ID NO:6. Geranylgeranyl pyrophosphate can be made by contacting farnesyl pyrophosphate and isopentenyl pyrophosphate with a polypeptide encoded by such nucleic acids. Expression vectors that  
25 include such nucleic acids operably linked to an expression control element also are featured.

Isolated nucleic acids having at least 82% sequence identity to the nucleotide sequence of SEQ ID NO:7 (e.g., at least 85%, 90%, or 95% sequence identity to the nucleotide sequence of SEQ ID NO:7) or to a fragment of SEQ ID NO:7 at least 30  
30 contiguous nucleotides in length also are featured. An isolated nucleic acid also can encode a phytoene desaturase polypeptide at least 90% identical to the amino acid

sequence of SEQ ID NO:8. Lycopene can be made by contacting phytoene with a polypeptide encoded by such nucleic acids. An expression vector that includes such nucleic acids operably linked to an expression control element also is featured.

The invention also features an isolated nucleic acid having at least 82% sequence identity to the nucleotide sequence of SEQ ID NO:9 (e.g., at least 85%, 90%, or 95% sequence identity to the nucleotide sequence of SEQ ID NO:9) or to a fragment of SEQ ID NO:9 at least 23 contiguous nucleotides in length. An isolated nucleic acid also can encode a phytoene synthase polypeptide at least 89% identical to the amino acid sequence of SEQ ID NO:10. Phytoene can be made by contacting geranylgeranyl pyrophosphate with a polypeptide encoded by such nucleic acids. An expression vector that includes such nucleic acids operably linked to an expression control element also is featured.

In yet another aspect, the invention features an isolated nucleic acid having at least 85% sequence identity to the nucleotide sequence of SEQ ID NO:11 (e.g., at least 90% or 95% identity to the nucleotide sequence of SEQ ID NO:11) or to a fragment of SEQ ID NO:11 at least 36 contiguous nucleotides in length. An isolated nucleic acid can encode a  $\beta$ -carotene hydroxylase polypeptide at least 90% identical to the amino acid sequence of SEQ ID NO:12. Zeaxanthin can be made by contacting  $\beta$ -carotene with a polypeptide encoded by such nucleic acids. Astaxanthin can be made by contacting canthaxanthin with a polypeptide encoded by such nucleic acids. The invention also features an expression vector that includes such nucleic acids operably linked to an expression control element.

The invention also features membranous bacteria (e.g., a *Rhodobacter* species) that include at least one exogenous nucleic acid encoding phytoene desaturase, lycopene  $\beta$ -cyclase,  $\beta$ -carotene hydroxylase, and  $\beta$ -carotene C4 oxygenase, wherein expression of the at least one exogenous nucleic acid produces detectable amounts of astaxanthin in the membranous bacteria. The amino acid sequence of the phytoene desaturase can be at least 90% identical to the amino acid sequence of SEQ ID NO:8. The amino acid sequence of the lycopene  $\beta$ -cyclase can be at least 83% identical to the amino acid sequence of SEQ ID NO:4. The amino acid sequence of the  $\beta$ -carotene hydroxylase can be at least 90% identical to the amino acid sequence of SEQ ID NO:12. The amino acid sequence of the  $\beta$ -carotene C4 oxygenase can be at least 80% identical to the amino acid

sequence of SEQ ID NO:39. The membranous bacteria further can include an exogenous nucleic acid encoding geranylgeranyl pyrophosphate synthase (e.g., a multifunctional geranylgeranyl pyrophosphate synthase) or can lack endogenous bacteriochlorophyll biosynthesis. The multifunctional geranylgeranyl pyrophosphate synthase can have an amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:45. The membranous bacteria further can include an exogenous nucleic acid encoding phytoene synthase. The phytoene synthase can have an amino acid sequence at least 89% identical to the amino acid sequence of SEQ ID NO:10.

In another aspect, the invention features membranous bacteria that include an exogenous nucleic acid encoding a phytoene desaturase having an amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:8, and wherein the membranous bacteria produces detectable amounts of lycopene. The membranous bacteria further can include a lycopene  $\beta$ -cyclase, wherein the membranous bacteria produce detectable amounts of  $\beta$ -carotene. The membranous bacteria also can include a  $\beta$ -carotene hydroxylase, wherein the membranous bacteria produce detectable amounts of zeaxanthin.

In still yet another aspect, the invention feature membranous bacteria that include at least one exogenous nucleic acid encoding phytoene desaturase, lycopene  $\beta$ -cyclase, and  $\beta$ -carotene C4 oxygenase, wherein expression of the at least one exogenous nucleic acid produces detectable amounts of canthaxanthin in the membranous bacteria. The membranous bacteria also can include a  $\beta$ -carotene hydroxylase, wherein the membranous bacteria produce detectable amounts of astaxanthin.

The invention also features a composition that includes an engineered *Rhodobacter* cell, wherein the cell produces a detectable amount of astaxanthin or canthaxanthin. The engineered *Rhodobacter* cell can include at least one exogenous nucleic acid encoding phytoene desaturase, lycopene  $\beta$ -cyclase,  $\beta$ -carotene hydroxylase, and  $\beta$ -carotene C4 oxygenase. The composition can be formulated for aquaculture and can pigment the flesh of fish or the carapace of crustaceans after ingestion. The composition can be formulated for human consumption or as an animal feed (e.g., formulated for consumption by chickens, turkeys, cattle, swine, or sheep).

The invention also features a method of making a nutraceutical. The method includes extracting carotenoids from an engineered *Rhodobacter* cell, the engineered *Rhodobacter* cell including at least one exogenous nucleic acid encoding phytoene desaturase, lycopene  $\beta$ -cyclase,  $\beta$ -carotene hydroxylase, and  $\beta$ -carotene C4 oxygenase, and wherein the *Rhodobacter* cell produces detectable amounts of astaxanthin.

In yet another aspect, the invention features membranous bacteria, wherein the membranous bacteria include an exogenous nucleic acid encoding a lycopene  $\beta$ -cyclase having an amino acid sequence at least 83% identical to the amino acid sequence of SEQ ID NO:4. The membranous bacteria further can include a phytoene desaturase, (e.g., an exogenous phytoene desaturase), wherein the membranous bacteria produce detectable amounts of  $\beta$ -carotene. The membranous bacteria also can include a  $\beta$ -carotene hydroxylase (e.g., an exogenous  $\beta$ -carotene hydroxylase), wherein the bacteria produce detectable amounts of zeaxanthin.

Membranous bacteria that include a  $\beta$ -carotene hydroxylase having an amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:12 also is featured. The membranous bacteria further can include a lycopene  $\beta$ -cyclase (e.g., an exogenous lycopene  $\beta$ -cyclase), wherein the membranous bacteria produce detectable amounts of zeaxanthin. The membranous bacteria also can include a phytoene desaturase (e.g., an exogenous phytoene desaturase), wherein the membranous bacteria produce detectable amounts of  $\beta$ -carotene.

The invention also features membranous bacteria (e.g., a *Rhodobacter* species) lacking an endogenous nucleic acid encoding a farnesyl pyrophosphate synthase, wherein the bacteria produces detectable amounts of carotenoids. The membranous bacteria also can include an exogenous nucleic acid encoding a multifunctional geranylgeranyl pyrophosphate synthase.

In another aspect, the invention features an isolated nucleic acid having at least 70% sequence identity (e.g., at least 80% or 90%) to the nucleotide sequences of SEQ ID NO:38, or to a fragment of the nucleic acid of SEQ ID NO:38 at least 15 contiguous nucleotides in length. The nucleic acid can encode a  $\beta$ -carotene C4 oxygenase. Canthaxanthin can be made by contacting  $\beta$ -carotene with a polypeptide encoded by such nucleic acids or a polypeptide having an amino acid sequence at least 80% identical to the

amino acid sequence of SEQ ID NO:39. Astaxanthin can be made by contacting zeaxanthin with a polypeptide encoded by such isolated nucleic acids or a polypeptide having an amino acid sequence at least 80% identical to the amino acid sequence of SEQ ID NO:39.

5           In another aspect, the invention features membranous bacteria that include an exogenous nucleic acid encoding a  $\beta$ -carotene C4 oxygenase, where the  $\beta$ -carotene oxygenase has an amino acid sequence at least 80% identical to the amino acid sequence of SEQ ID NO:39.

10           In yet another aspect, the invention features a host cell comprising an exogenous nucleic acid, wherein the exogenous nucleic acid includes a nucleic acid sequence encoding one or more polypeptides that catalyze the formation of (3S, 3'S) astaxanthin, wherein the host cell produces CoQ-10 and (3S, 3'S) astaxanthin. A method of making CoQ-10 and (3S, 3'S) astaxanthin at substantially the same time also is featured. The method includes transforming a host cell with a nucleic acid, wherein the nucleic acid  
15 includes a nucleic acid sequence that encodes one or more polypeptides, wherein the polypeptides catalyze the formation of (3S, 3'S) astaxanthin; and culturing the host cell under conditions that allow for the production of (3S, 3'S) astaxanthin and CoQ-10. The method further can include transforming the host cell with at least one exogenous nucleic acid, the exogenous nucleic acid encoding one or more polypeptides, wherein the  
20 polypeptides catalyze the formation of CoQ-10.

The invention also features isolated nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:38, and SEQ ID NO:44.

25           An isolated nucleic acid having at least 90% sequence identity to the nucleotide sequences of SEQ ID NO:44, or to a fragment of the nucleic acid of SEQ ID NO:44 at least 60 contiguous nucleotides in length is featured. Geranylgeranyl pyrophosphate can be made by contacting isopentenyl pyrophosphate and dimethylallyl pyrophosphate with a polypeptide encoded by such a nucleic acid.

30           Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those

described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

### DESCRIPTION OF DRAWINGS

FIG 1 is a schematic diagram of the biosynthetic pathway for the production of zeaxanthin and conversion to zeaxanthin di-glucoside.

FIG 2 is a schematic diagram of the *P. stewartii* carotenoid gene operon (6586 bp).

FIG 3 is a chromatogram of astaxanthin production in *P. stewartii::crtW(B. aurantiaca)*.

### DETAILED DESCRIPTION

#### Nucleic Acid Molecules

The invention features isolated nucleic acids that encode enzymes involved in carotenoid biosynthesis. The nucleic acids of SEQ ID NO:1, 3, 5, 7, 9, and 11 encode zeaxanthin glucosyl transferase (*crtX*), lycopene  $\beta$ -cyclase (*crtY*), geranylgeranyl-pyrophosphate synthase (*crtE*), phytoene desaturase (*crtI*), phytoene synthase (*crtB*) and  $\beta$ -carotene hydroxylase (*crtZ*), respectively. A nucleic acid of the invention can have at least 76% sequence identity, e.g., 78%, 80%, 85%, 90%, 95%, or 99% sequence identity, to the nucleic acid of SEQ ID NO:1, or to fragments of the nucleic acid of SEQ ID NO:1 that are at least about 33 nucleotides in length; at least 78% sequence identity, e.g., 80%, 85%, 90%, 95%, or 99% sequence identity, to the nucleotide sequence of SEQ ID NO:3, or to fragments of the nucleic acid of SEQ ID NO:3 that are at least about 32 nucleotides in length; at least 81% sequence identity, e.g., 82%, 85%, 90%, 95%, or 99% sequence identity, to the nucleotide sequence of SEQ ID NO:5, or to fragments of the nucleic acid of SEQ ID NO:5 that are at least about 60 nucleotides in length; at least 82% sequence identity, e.g., 83%, 85%, 90%, 95%, or 99% sequence identity, to the nucleotide



sequences of SEQ ID NO:7 or SEQ ID NO:9, or to fragments of the nucleic acids of SEQ ID NO:7 or SEQ ID NO:9 that are at least about 30 or 23 nucleotides in length, respectively; at least 85% sequence identity, e.g., 86%, 90%, 92%, 95%, or 99% sequence identity, to the nucleotide sequence of SEQ ID NO:11, or to fragments of the nucleic acid of SEQ ID NO:11 that are at least about 36 nucleotides in length. A nucleic acid of the invention can have at least 60% sequence identity, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% sequence identity to the nucleotide sequence of SEQ ID NO:38 or to fragments of the nucleic acid of SEQ ID NO:38 that are at least about 15 nucleotides in length. Such a nucleic acid can encode a  $\beta$ -carotene C4 oxygenase (*crtW*). A nucleic acid of the invention also can have at least 90% identity to the nucleotide sequence set forth in SEQ ID NO:44 or to fragments of the nucleic acid of SEQ ID NO:44 that are at least about 60 nucleotides in length. Such a nucleic acid can encode a multifunctional geranylgeranyl pyrophosphate synthase.

Generally, percent sequence identity is calculated by determining the number of matched positions in aligned nucleic acid sequences, dividing the number of matched positions by the total number of aligned nucleotides, and multiplying by 100. A matched position refers to a position in which identical nucleotides occur at the same position in aligned nucleic acid sequences. Percent sequence identity can be determined for any nucleic acid or amino acid sequence as follows. First, a nucleic acid or amino acid sequence is compared to the identified nucleic acid or amino acid sequence using the BLAST 2 Sequences (B12seq) program from the stand-alone version of BLASTZ containing BLASTN version 2.0.14 and BLASTP version 2.0.14. This stand-alone version of BLASTZ can be obtained from the University of Wisconsin library as well as at [www.fr.com](http://www.fr.com) or [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). Instructions explaining how to use the B12seq program can be found in the readme file accompanying BLASTZ.

B12seq performs a comparison between two sequences using either the BLASTN or BLASTP algorithm. BLASTN is used to compare nucleic acid sequences, while BLASTP is used to compare amino acid sequences. To compare two nucleic acid sequences, the options are set as follows: -i is set to a file containing the first nucleic acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second nucleic acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastn; -o is set to any

desired file name (e.g., C:\output.txt); -q is set to -1; -r is set to 2; and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two sequences: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastn -o c:\output.txt -q -1 -r 2. To compare two amino acid sequences, the options of Bl2seq are set as follows: -i is set to a file containing the first amino acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second amino acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastp; -o is set to any desired file name (e.g., C:\output.txt); and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two amino acid sequences: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastp -o c:\output.txt. If the target sequence shares homology with any portion of the identified sequence, then the designated output file will present those regions of homology as aligned sequences. If the target sequence does not share homology with any portion of the identified sequence, then the designated output file will not present aligned sequences.

Once aligned, a length is determined by counting the number of consecutive nucleotides or amino acid residues from the target sequence presented in alignment with sequence from the identified sequence starting with any matched position and ending with any other matched position. A matched position is any position where an identical nucleotide or amino acid residue is presented in both the target and identified sequence. Gaps presented in the target sequence are not counted since gaps are not nucleotides or amino acid residues. Likewise, gaps presented in the identified sequence are not counted since target sequence nucleotides or amino acid residues are counted, not nucleotides or amino acid residues from the identified sequence.

The percent identity over a particular length is determined by counting the number of matched positions over that length and dividing that number by the length followed by multiplying the resulting value by 100. For example, if (1) a 1000 nucleotide target sequence is compared to the sequence set forth in SEQ ID NO:1, (2) the Bl2seq program presents 200 nucleotides from the target sequence aligned with a region of the sequence set forth in SEQ ID NO: 1 where the first and last nucleotides of that 200 nucleotide region are matches, and (3) the number of matches over those 200 aligned nucleotides is

180, then the 1000 nucleotide target sequence contains a length of 200 and a percent identity over that length of 90 (i.e.  $180 \div 200 * 100 = 90$ ).

It will be appreciated that a single nucleic acid or amino acid target sequence that aligns with an identified sequence can have many different lengths with each length having its own percent identity. For example, a target sequence containing a 20 nucleotide region that aligns with an identified sequence as follows has many different lengths including those listed in Table 1.

	1	20
Target Sequence:	AGGTCGTGTACTGTCAGTCA (SEQ ID NO:46)	
Identified Sequence:	ACGTGGTGAAGTCCAGTGA (SEQ ID NO:47)	

**TABLE 1**

Starting Position	Ending Position	Length	Matched Positions	Percent Identity
1	20	20	15	75.0
1	18	18	14	77.8
1	15	15	11	73.3
6	20	15	12	80.0
6	17	12	10	83.3
6	15	10	8	80.0
8	20	13	10	76.9
8	16	9	7	77.8

It is noted that the percent identity value is rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 is rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 is rounded up to 78.2. It is also noted that the length value will always be an integer.

Isolated nucleic acid molecules of the invention are at least about 20 nucleotides in length. For example, the nucleic acid molecule can be about 20-30, 22-32, 33-50, 34 to 45, 40-50, 60-80, 62 to 92, 50-100, or greater than 150 nucleotides in length, e.g., 200-300, 300-500, or 500-1000 nucleotides in length. Such fragments, whether protein-encoding or not, can be used as probes, primers, and diagnostic reagents. In some embodiments, the isolated nucleic acid molecules encode a full-length zeaxanthin glucosyl transferase, lycopene  $\beta$ -cyclase, geranylgeranyl pyrophosphate synthase, phytoene desaturase,  $\beta$ -carotene hydroxylase,  $\beta$ -carotene C4 oxygenase, or

multifunctional geranylgeranyl pyrophosphate synthase polypeptide. Nucleic acid molecules can be DNA or RNA, linear or circular, and in sense or antisense orientation.

Isolated nucleic acid molecules of the invention can be produced by standard techniques. As used herein, "isolated" refers to a sequence corresponding to part or all of  
5 a gene encoding a zeaxanthin glucosyl transferase, lycopene  $\beta$ -cyclase, geranylgeranyl-pyrophosphate synthase, phytoene desaturase, phytoene synthase,  $\beta$ -carotene hydroxylase,  $\beta$ -carotene C4 oxygenase, or multifunctional geranylgeranyl pyrophosphate synthase polypeptide, or an operon encoding two or more such polypeptides, but free of sequences that normally flank one or both sides of the wild-type gene or the operon in a  
10 naturally-occurring genome, e.g., a bacterial genome. The term "isolated" as used herein with respect to nucleic acids also includes any non-naturally-occurring nucleic acid sequence since such non-naturally-occurring sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome.

An isolated nucleic acid can be, for example, a DNA molecule, provided one of  
15 the nucleic acid sequences normally found immediately flanking that DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a DNA molecule that exists as a separate molecule (e.g., a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as recombinant DNA that is  
20 incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a retrovirus, adenovirus, or herpes virus), or into the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include an engineered nucleic acid such as a recombinant DNA molecule that is part of a hybrid or fusion nucleic acid. A nucleic acid existing among hundreds to millions of other nucleic acids within, for  
25 example, cDNA libraries or genomic libraries, or gel slices containing a genomic DNA restriction digest, is not to be considered an isolated nucleic acid.

Isolated nucleic acids within the scope of the invention can be obtained using any method including, without limitation, common molecular cloning and chemical nucleic acid synthesis techniques. For example, polymerase chain reaction (PCR) techniques can  
30 be used to obtain an isolated nucleic acid containing a nucleic acid sequence sharing identity with the sequences set forth in SEQ ID NOs: 1, 3, 5, 7, 9, 11, 38, or 44. PCR

refers to a procedure or technique in which target nucleic acids are amplified. Sequence information from the ends of the region of interest or beyond typically is employed to design oligonucleotide primers that are identical in sequence to opposite strands of the template to be amplified. PCR can be used to amplify specific sequences from DNA as well as RNA, including sequences from total genomic DNA or total cellular RNA.

Primers are typically 14 to 40 nucleotides in length, but can range from 10 nucleotides to hundreds of nucleotides in length. General PCR techniques are described, for example in PCR Primer: A Laboratory Manual, Ed. by Dieffenbach, C. and Dveksler, G., Cold Spring Harbor Laboratory Press, 1995. When using RNA as a source of template, reverse transcriptase can be used to synthesize complementary DNA (cDNA) strands.

Isolated nucleic acids of the invention also can be chemically synthesized, either as a single nucleic acid molecule or as a series of oligonucleotides. For example, one or more pairs of long oligonucleotides (e.g., >100 nucleotides) can be synthesized that contain the desired sequence, with each pair containing a short segment of complementary (e.g., about 15 nucleotides) DNA such that a duplex is formed when the oligonucleotide pair is annealed. DNA polymerase is used to extend the oligonucleotides, resulting in a double-stranded nucleic acid molecule per oligonucleotide pair, which then can be ligated into a vector.

Isolated nucleic acids of the invention also can be obtained by mutagenesis. For example, an isolated nucleic acid that shares identity with a sequence set forth in SEQ ID NO: 1, 3, 5, 7, 9, 11, 38, or 44 can be mutated using common molecular cloning techniques (e.g., site-directed mutagenesis). Possible mutations include, without limitation, deletions, insertions, and substitutions, as well as combinations of deletions, insertions, and substitutions. Alignments of nucleic acids of the invention with other known sequences encoding carotenoid enzymes can be used to identify positions to modify. For example, alignment of the nucleotide sequence of SEQ ID NO:5 with other nucleic acids encoding geranyl geranyl pyrophosphate synthases (e.g., from *Erwinia uredovora*) provides guidance as to which nucleotides can be substituted, which nucleotides can be deleted, and at which positions nucleotides can be inserted.

In addition, nucleic acid and amino acid databases (e.g., GenBank<sup>®</sup>) can be used to obtain an isolated nucleic acid within the scope of the invention. For example, any

nucleic acid sequence having homology to a sequence set forth in SEQ ID NO: 1, 3, 5, 7, 9, 11, 38, or 44, or any amino acid sequence having homology to a sequence set forth in SEQ ID NO: 2, 4, 6, 8, 10, 12, 39, or 45 can be used as a query to search GenBank®.

Furthermore, nucleic acid hybridization techniques can be used to obtain an isolated nucleic acid within the scope of the invention. Briefly, any nucleic acid having some homology to a sequence set forth in SEQ ID NO: 1, 3, 5, 7, 9, 11, 38, or 44 can be used as a probe to identify a similar nucleic acid by hybridization under conditions of moderate to high stringency. Moderately stringent hybridization conditions include hybridization at about 42°C in a hybridization solution containing 25 mM KPO<sub>4</sub> (pH 7.4), 5X SSC, 5X Denhart's solution, 50 µg/mL denatured, sonicated salmon sperm DNA, 50% formamide, 10% Dextran sulfate, and 1-15 ng/mL probe (about 5x10<sup>7</sup> cpm/µg), and wash steps at about 50°C with a wash solution containing 2X SSC and 0.1% SDS. For high stringency, the same hybridization conditions can be used, but washes are performed at about 65°C with a wash solution containing 0.2X SSC and 0.1% SDS.

Once a nucleic acid is identified, the nucleic acid then can be purified, sequenced, and analyzed to determine whether it is within the scope of the invention as described herein. Hybridization can be done by Southern or Northern analysis to identify a DNA or RNA sequence, respectively, that hybridizes to a probe. The probe can be labeled with biotin, digoxigenin, an enzyme, or a radioisotope such as <sup>32</sup>P or <sup>35</sup>S. The DNA or RNA to be analyzed can be electrophoretically separated on an agarose or polyacrylamide gel, transferred to nitrocellulose, nylon, or other suitable membrane, and hybridized with the probe using standard techniques well known in the art. See, for example, sections 7.39-7.52 of Sambrook *et al.*, (1989) Molecular Cloning, second edition, Cold Spring harbor Laboratory, Plainview, NY.

## Polypeptides

The present invention also features isolated zeaxanthin glucosyl transferase (SEQ ID NO:2), lycopene β-cyclase (SEQ ID NO:4), geranylgeranyl pyrophosphate synthase (SEQ ID NO:6), phytoene desaturase (SEQ ID NO:8), phytoene synthase (SEQ ID NO:10), and β-carotene hydroxylase (SEQ ID NO:12) polypeptides. In addition, the invention features isolated β-carotene C4 oxygenase polypeptides (SEQ ID NO:39) and

multifunctional geranylgeranyl pyrophosphate synthase polypeptides (SEQ ID NO:45). A polypeptide of the invention can have at least 75% sequence identity, e.g., 80%, 85%, 90%, 95%, or 99% sequence identity, to the amino acid sequence of SEQ ID NO:2 or to fragments thereof; at least 83% sequence identity, e.g., 85%, 90%, 95%, or 99% sequence identity, to the amino acid sequence of SEQ ID NO:4 or to fragments thereof; at least 85% sequence identity, e.g., 90%, 95%, or 99% sequence identity, to the amino acid sequence of SEQ ID NO:6 or to fragments thereof; at least 90% sequence identity, e.g., 90%, 92%, 95%, or 99% sequence identity, to the amino acid sequence of SEQ ID NO:8 or to fragments thereof; at least 89% sequence identity, e.g., 90%, 95%, or 99% sequence identity, to the amino acid sequence of SEQ ID NO:10 or to fragments thereof; at least 90% sequence identity, e.g., 95%, or 99% sequence identity, to the amino acid sequence of SEQ ID NO:12 or to fragments thereof; at least 60% sequence identity, e.g., 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% sequence identity, to the amino acid sequence of SEQ ID NO:39 or to fragments thereof; or at least 90% sequence identity, e.g., 95% or 99% sequence identity, to the amino acid sequence set forth in SEQ ID NO:45 or to fragments thereof. Percent sequence identity can be determined as described above for nucleic acid molecules.

An "isolated polypeptide" has been separated from cellular components that naturally accompany it. Typically, the polypeptide is isolated when it is at least 60% (e.g., 70%, 80%, 90%, 95%, or 99%), by weight, free from proteins and naturally-occurring organic molecules that are naturally associated with it. In general, an isolated polypeptide will yield a single major band on a non-reducing polyacrylamide gel.

The term "polypeptide" includes any chain of amino acids, regardless of length or post-translational modification. Polypeptides that have identity to the amino acid sequences of SEQ ID NO:2, 4, 6, 8, 10, 12, 39, or 45 can retain the function of the enzyme (see FIG 1 for a schematic of the carotenoid biosynthesis pathway). For example, geranylgeranyl pyrophosphate synthase can produce geranylgeranyl pyrophosphate (GGPP) by condensing together isopentenyl pyrophosphate (IPP) with farnesyl pyrophosphate (FPP). Phytoene synthase can produce phytoene by condensing together two molecules of GGPP. Phytoene desaturase can perform four successive desaturations on phytoene to form lycopene. Lycopene  $\beta$ -cyclase can perform two

successive cyclization reactions on lycopene to form  $\beta$ -carotene.  $\beta$ -carotene hydroxylase can perform two successive hydroxylation reactions on  $\beta$ -carotene to form zeaxanthin. Alternatively,  $\beta$ -carotene hydroxylase can perform two successive hydroxylation reactions on canthaxanthin to form astaxanthin. Zeaxanthin glucosyl transferase can add one or two glucose or other sugar moieties to zeaxanthin to form zeaxanthin monoglycoside or diglycoside, respectively.  $\beta$ -carotene C4 oxygenase can convert the methylene groups at the C4 and C4' positions of the  $\beta$ -carotene or zeaxanthin to form canthaxanthin or astaxanthin, respectively. Multifunctional geranylgeranyl pyrophosphate synthase can directly convert 3 IPP molecules and 1 dimethylallyl pyrophosphate (DMAPP) molecule to 1 GGPP molecule.

In general, conservative amino acid substitutions, i.e., substitutions of similar amino acids, are tolerated without affecting protein function. Similar amino acids are those that are similar in size and/or charge properties. Families of amino acids with similar side chains are known. These families include amino acids with basic side chains (e.g., lysine, arginine, or histidine), acidic side chains (e.g., aspartic acid or glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, or cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, or tryptophan),  $\beta$ -branched side chains (e.g., threonine, valine, or isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, or histidine).

Mutagenesis also can be used to alter a nucleic acid such that activity of the polypeptide encoded by the nucleic acid is altered (e.g., to increase production of a particular carotenoid). For example, error-prone PCR (e.g., (GeneMorph PCR Mutagenesis Kit; Stratagene Inc. La Jolla, CA; Catalog # 600550; Revision #090001) can be used to mutagenize the *B. aurantiaca crtW* gene (SEQ ID NO:38) to increase the relative amount of di-keto carotenoid (e.g. astaxanthin (3,3'-dihydroxy- $\beta,\beta$ -carotene-4,4'-dione) or canthaxanthin ( $\beta,\beta$ -carotene-4,4'-dione)) relative to mono-keto carotenoid (e.g. echinone ( $\beta,\beta$ -carotene-4-one) or adonixanthin (3,3'-dihydroxy- $\beta,\beta$ -carotene-4-one)) that is produced. In general, the nucleic acid to be mutagenized can be cloned into a vector such as pCR-Blunt II-TOPO (Clontech; Palo Alto, CA) and used as a template for error-prone PCR. For purposes of directed evolution, mutation frequencies of 2-7 nucleotides /



Kbp template (1-4 amino acids mutations / 333 Amino acids) generally are desired. Mutation frequency can be lowered or raised by increasing or decreasing the template concentration, respectively. PCR can be performed according to manufacturer's recommendations. Mutagenized nucleic acid is ligated into an expression vector, which is  
5 used to transform a host, and activity of the expressed protein is assessed. For example, in the case of the *crtW* gene, electrocompetent *P. stewartii* (ATCC 8200) cells can be prepared and transformed as described herein, and resulting individual colonies can be screened by visual inspection for a phenotypic change from bright yellow pigmentation (production of zeaxanthin), yellow orange (production of mono-keto carotenoid) or  
10 reddish-orange (production of di-keto carotenoid). Production of increased amounts of astaxanthin can be confirmed by HPLC/MS.

Isolated polypeptides of the invention can be obtained, for example, by extraction from a natural source (e.g., a plant or bacteria cell), chemical synthesis, or by recombinant production in a host. For example, a polypeptide of the invention can be  
15 produced by ligating a nucleic acid molecule encoding the polypeptide into a nucleic acid construct such as an expression vector, and transforming a bacterial or eukaryotic host cell with the expression vector. In general, nucleic acid constructs include expression control elements operably linked to a nucleic acid sequence encoding a polypeptide of the invention (e.g., zeaxanthin glucosyl transferase, lycopene  $\beta$ -cyclase, geranylgeranyl  
20 pyrophosphate synthase, phytoene desaturase, phytoene synthase,  $\beta$ -carotene hydroxylase,  $\beta$ -carotene C4 oxygenase, or multifunctional geranylgeranyl pyrophosphate synthase polypeptides). Expression control elements do not typically encode a gene product, but instead affect the expression of the nucleic acid sequence. As used herein, "operably linked" refers to connection of the expression control elements to the nucleic  
25 acid sequence in such a way as to permit expression of the nucleic acid sequence. Expression control elements can include, for example, promoter sequences, enhancer sequences, response elements, polyadenylation sites, or inducible elements. Non-limiting examples of promoters include the *puf* promoter from *Rhodobacter sphaeroides* (GenBank Accession No. E13945), the *nifHDK* promoter from *R. sphaeroides* (GenBank  
30 Accession No. AF031817), and the *fliK* promoter from *R. sphaeroides* (GenBank Accession No. U86454).

In bacterial systems, a strain of *E. coli* such as DH10B or BL-21 can be used. Suitable *E. coli* vectors include, but are not limited to, pUC18, pUC19, the pGEX series of vectors that produce fusion proteins with glutathione S-transferase (GST), and pBluescript series of vectors. Transformed *E. coli* are typically grown exponentially then stimulated with isopropylthiogalactopyranoside (IPTG) prior to harvesting. In general, fusion proteins produced from the pGEX series of vectors are soluble and can be purified easily from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites such that the cloned target gene product can be released from the GST moiety.

In eukaryotic host cells, a number of viral-based expression systems can be utilized to express polypeptides of the invention. A nucleic acid encoding a polypeptide of the invention can be cloned into, for example, a baculoviral vector such as pBlueBac (Invitrogen, San Diego, CA) and then used to co-transfect insect cells such as *Spodoptera frugiperda* (Sf9) cells with wild-type DNA from *Autographa californica* multiply enveloped nuclear polyhedrosis virus (AcMNPV). Recombinant viruses producing polypeptides of the invention can be identified by standard methodology. Alternatively, a nucleic acid encoding a polypeptide of the invention can be introduced into a SV40, retroviral, or vaccinia based viral vector and used to infect suitable host cells.

A polypeptide within the scope of the invention can be "engineered" to contain an amino acid sequence that allows the polypeptide to be captured onto an affinity matrix. For example, a tag such as c-myc, hemagglutinin, polyhistidine, or Flag<sup>TM</sup> tag (Kodak) can be used to aid polypeptide purification. Such tags can be inserted anywhere within the polypeptide including at either the carboxyl or amino termini. Other fusions that could be useful include enzymes that aid in the detection of the polypeptide, such as alkaline phosphatase.

*Agrobacterium*-mediated transformation, electroporation and particle gun transformation can be used to transform plant cells. Illustrative examples of transformation techniques are described in U.S. Patent No. 5,204,253 (particle gun) and U.S. Patent No. 5,188,958 (*Agrobacterium*). Transformation methods utilizing the Ti and Ri plasmids of *Agrobacterium spp.* typically use binary type vectors. Walkerpeach, C. et

al., in Plant Molecular Biology Manual, S. Gelvin and R. Schilperoort, eds., Kluwer Dordrecht, C1:1-19 (1994). If cell or tissue cultures are used as the recipient tissue for transformation, plants can be regenerated from transformed cultures by techniques known to those skilled in the art.

5

#### *Engineered cells*

Any cell containing an isolated nucleic acid within the scope of the invention is itself within the scope of the invention. This includes, without limitation, prokaryotic cells such as *R. sphaeroides* cells and eukaryotic cells such as plant, yeast, and other  
10 fungal cells. It is noted that cells containing an isolated nucleic acid of the invention are not required to express the isolated nucleic acid. In addition, the isolated nucleic acid can be integrated into the genome of the cell or maintained in an episomal state. In other words, cells can be stably or transiently transfected with an isolated nucleic acid of the invention.

15

Any method can be used to introduce an isolated nucleic acid into a cell. In fact, many methods for introducing nucleic acid into a cell, whether *in vivo* or *in vitro*, are well known to those skilled in the art. For example, calcium phosphate precipitation, conjugation, electroporation, heat shock, lipofection, microinjection, and viral-mediated nucleic acid transfer are common methods that can be used to introduce nucleic acid  
20 molecules into a cell. In addition, naked DNA can be delivered directly to cells *in vivo* as describe elsewhere (U.S. Patent Nos. 5,580,859 and 5,589,466). Furthermore, nucleic acid can be introduced into cells by generating transgenic animals.

Any method can be used to identify cells that contain an isolated nucleic acid within the scope of the invention. For example, PCR and nucleic acid hybridization  
25 techniques such as Northern and Southern analysis can be used. In some cases, immunohistochemistry and biochemical techniques can be used to determine if a cell contains a particular nucleic acid by detecting the expression of a polypeptide encoded by that particular nucleic acid. For example, the polypeptide of interest can be detected with an antibody having specific binding affinity for that polypeptide, which indicates that that  
30 cell not only contains the introduced nucleic acid but also expresses the encoded polypeptide. Enzymatic activities of the polypeptide of interest also can be detected or an

end product (e.g., a particular carotenoid) can be detected as an indication that the cell contains the introduced nucleic acid and expresses the encoded polypeptide from that introduced nucleic acid.

5 The cells described herein can contain a single copy, or multiple copies (e.g., about 5, 10, 20, 35, 50, 75, 100 or 150 copies), of a particular exogenous nucleic acid. All non-naturally-occurring nucleic acids are considered an exogenous nucleic acid once introduced into the cell. The term "exogenous" as used herein with reference to a nucleic acid and a particular cell refers to any nucleic acid that does not originate from that particular cell as found in nature. Nucleic acid that is naturally-occurring also can be  
10 exogenous to a particular cell. For example, an entire operon that is isolated from a bacteria is an exogenous nucleic acid with respect to a second bacteria once that operon is introduced into the second bacteria. For example, a bacterial cell (e.g., *Rhodobacter*) can contain about 50 copies of an exogenous nucleic acid of the invention. In addition, the cells described herein can contain more than one particular exogenous nucleic acid. For  
15 example, a bacterial cell can contain about 50 copies of exogenous nucleic acid X as well as about 75 copies of exogenous nucleic acid Y. In these cases, each different nucleic acid can encode a different polypeptide having its own unique enzymatic activity. For example, a bacterial cell can contain two different exogenous nucleic acids such that a high level of astaxanthin or other carotenoid is produced. In addition, a single exogenous  
20 nucleic acid can encode one or more polypeptides. For example, a single nucleic acid can contain sequences that encode three or more different polypeptides.

Microorganisms that are suitable for producing carotenoids may or may not naturally produce carotenoids, and include prokaryotic and eukaryotic microorganisms, such as bacteria, yeast, and fungi. In particular, yeast such as *Phaffia rhodozyma*  
25 (*Xanthophyllomyces dendrorhous*), *Candida utilis*, and *Saccharomyces cerevisiae*, fungi such as *Neurospora crassa*, *Phycomyces blakesleeanus*, *Blakeslea trispora*, and *Aspergillus sp.*, Archaeobacteria such as *Halobacterium salinarium*, and Eubacteria including *Pantoea* species (formerly called *Erwinia*) such as *Pantoea stewartii* (e.g., ATCC Accession #8200), flavobacteria species such as *Xanthobacter autotrophicus* and  
30 *Flavobacterium multivorum*, *Zymomonas mobilis*, *Rhodobacter* species such as *R. sphaeroides* and *R. capsulatus*, *E. coli*, and *E. vulneris* can be used. Other examples of

bacteria that may be used include bacteria in the genus *Sphingomonas* and Gram negative bacteria in the  $\alpha$ -subdivision, including, for example, *Paracoccus*, *Azotobacter*, *Agrobacterium*, and *Erythrobacter*. Eubacteria, and especially *R. sphaeroides* and *R. capsulatus*, are particularly useful. *R. sphaeroides* and *R. capsulatus* naturally produce certain carotenoids and grows on defined media. Such *Rhodobacter* species also are non-pyrogenic, minimizing health concerns about use in nutritional supplements. In some embodiments, it can be useful to produce carotenoids in plants and algae such as *Zea mays*, *Brassica napus*, *Lycopersicon esculentum*, *Tagetes erecta*, *Haematococcus pluvialis*, *Dunaliella salina*, *Chlorella protothecoides*, and *Neosporangiococcum excentrum*.

It is noted that bacteria can be membranous or non-membranous bacteria. The term "membranous bacteria" as used herein refers to any naturally-occurring, genetically modified, or environmentally modified bacteria having an intracytoplasmic membrane. An intracytoplasmic membrane can be organized in a variety of ways including, without limitation, vesicles, tubules, thylakoid-like membrane sacs, and highly organized membrane stacks. Any method can be used to analyze bacteria for the presence of intracytoplasmic membranes including, without limitation, electron microscopy, light microscopy, and density gradients. See, e.g., Chory et al., (1984) *J. Bacteriol.*, 159:540-554; Niederman and Gibson, Isolation and Physiochemical Properties of Membranes from Purple Photosynthetic Bacteria. In: The Photosynthetic Bacteria, Ed. By Roderick K. Clayton and William R. Sistrom, Plenum Press, pp. 79-118 (1978); and Lueking et al., (1978) *J. Biol. Chem.*, 253: 451-457. Examples of membranous bacteria that can be used include, without limitation, Purple Non-Sulfur Bacteria, including bacteria of the Rhodospirillaceae family such as those in the genus *Rhodobacter* (e.g., *R. sphaeroides* and *R. capsulatus*), the genus *Rhodospirillum*, the genus *Rhodopseudomonas*, the genus *Rhodomicrobium*, and the genus *Rhodophila*. The term "non-membranous bacteria" refers to any bacteria lacking intracytoplasmic membrane. Membranous bacteria can be highly membranous bacteria. The term "highly membranous bacteria" as used herein refers to any bacterium having more intracytoplasmic membrane than *R. sphaeroides* (ATCC 17023) cells have after the *R. sphaeroides* (ATCC 17023) cells have been (1) cultured chemoheterotrophically under aerobic condition for four days, (2) cultured

chemoheterotrophically under anaerobic for four hours, and (3) harvested. Aerobic culture conditions include culturing the cells in the dark at 30°C in the presence of 25% oxygen. Anaerobic culture conditions include culturing the cells in the light at 30°C in the presence of 2% oxygen. After the four hour anaerobic culturing step, the *R. sphaeroides* (ATCC 17023) cells are harvested by centrifugation and analyzed.

Nucleic acids of the invention can be expressed in microorganisms so that detectable amounts of carotenoids are produced. As used herein, "detectable" refers to the ability to detect the carotenoid and any esters or glycosides thereof using standard analytical methodology. In general, carotenoids can be extracted with an organic solvent such as acetone or methanol and detected by an absorption scan from 400-500 nm in the same organic solvent. In some cases, it is desirable to back-extract with a second organic solvent, such as hexane. The maximal absorbance of each carotenoid depends on the solvent that it is in. For example, in acetone, the maximal absorbance of lutein is at 451 nm, while maximal absorbance of zeaxanthin is at 454 nm. In hexane, the maximal absorbance of lutein and zeaxanthin is 446 nm and 450 nm, respectively. High performance liquid chromatography coupled to mass spectrometry also can be used to detect carotenoids. Two reverse phase columns that are connected in series can be used with a solvent gradient of water and acetone. The first column can be a C30 specialty column designed for carotenoid separation (e.g., YMCä Carotenoid S3m; 2.0 x 150 mm, 3mm particle size; Waters Corporation, PN CT99S031502WT) followed by a C8 Xterraä MS column (e.g., Xterraä MS C8; 2.1 x 250 mm, 5mm particle size; Waters Corporation, PN 186000459).

Detectable amounts of carotenoids include 10µg/g dry cell weight (dcw) and greater. For example, about 10 to 100,000µg/g dcw, about 100 to 60,000µg/g dcw, about 500 to 30,000µg/g dcw, about 1000 to 20,000 µg/g dcw, about 5,000 to 55,000 µg/g dcw, or about 30,000 µg/g dcw to about 55,000 µg/g dcw. With respect to algae or other plants or organisms that produce a particular carotenoid, such as astaxanthin, β-carotene, lycopene, or zeaxanthin, "detectable amount" of carotenoid is an amount that is detectable over the endogenous level in the plant or organism.

Depending on the microorganism and the metabolites present within the microorganism, one or more of the following enzymes may be expressed in the

microorganism: geranylgeranyl pyrophosphate synthase, phytoene synthase, phytoene desaturase, lycopene  $\beta$  cyclase, lycopene  $\epsilon$  cyclase, zeaxanthin glycosyl transferase,  $\beta$ -carotene hydroxylase,  $\beta$ -carotene C-4 ketolase, and multifunctional geranylgeranyl pyrophosphate synthase. Suitable nucleic acids encoding these enzymes are described  
5 above. Also, see, for example, Genbank Accession No. Y15112 for the sequence of carotenoid biosynthesis genes of *Paracoccus marcusii*; Genbank Accession No. D58420 for the carotenoid biosynthesis genes of *Agrobacterium aurantiacum*; Genbank Accession No. M87280 M99707 for the sequence of carotenoid biosynthesis genes of *Erwinia herbicola*; and Genbank Accession No. U62808 for carotenoid biosynthesis genes of  
10 *Flavobacterium* sp. Strain R1534.

For example, to produce lycopene in a microorganism that naturally produces neurosporene, such as *Rhodobacter*, an exogenous nucleic acid encoding phytoene desaturase can be expressed, e.g., a phytoene desaturase of the invention, and lycopene can be detected using standard methodology. Expression of additional carotenoid genes  
15 in such an engineered cell will allow for production of additional carotenoids. For example, expression of a lycopene  $\beta$ -cyclase in such an engineered cell allows production of detectable amounts of  $\beta$ -carotene, while further expression of a  $\beta$ -carotene hydroxylase allows production of another carotenoid, zeaxanthin.  $\beta$ -carotene and zeaxanthin can be detected using standard methodology and are distinguished by mobility on an HPLC  
20 column. Zeaxanthin diglucoside can be produced by further expression of zeaxanthin glucosyl transferase (*crtX*) in an organism that produces zeaxanthin.

Alternatively, canthaxanthin can be produced in organisms that produce phytoene by expression of phytoene desaturase, lycopene  $\beta$ -cyclase, and  $\beta$ -carotene C4 oxygenase, an enzyme that converts the methylene groups at the C4 and C4' positions of the  
25 carotenoid to ketone groups. The  $\beta$ -carotene C4 oxygenase from, e.g., *Agrobacterium aurantiacum* or *Haematococcus pluvialis* can be used. See, GenBank Accession Nos. 1136630 and X86782 for a description of the nucleotide and amino acid sequences of the *A. aurantiacum* and *H. pluvialis* enzymes, respectively. The  $\beta$ -carotene C4 oxygenase from *Brevundimonas aurantiaca* also can be used. See, Example 2 for a description of  
30 the nucleotide and amino acid sequences. In organisms that do not naturally produce carotenoids, additional enzymes are required for production of canthaxanthin.

Geranylgeranyl pyrophosphate synthase and phytoene synthase can be expressed such that the necessary precursors for canthaxanthin synthesis are present.

Astaxanthin also can be produced in microorganisms that naturally produce carotenoids. For example, a *Rhodobacter* cell can be engineered such that phytoene desaturase, lycopene  $\beta$ -cyclase,  $\beta$ -carotene hydroxylase, and  $\beta$ -carotene C4 oxygenase are expressed and detectable amounts of astaxanthin are produced. Such an organism also can express an enzyme that can modify the 3 or 3' hydroxyl groups of astaxanthin with chemical groups such as glucose (e.g., to produce astaxanthin diglucoside), other sugars, or fatty acids. In addition, a *P. stewartii* cell can be engineered such that  $\beta$ -carotene C4 oxygenase is expressed and detectable amounts of astaxanthin are produced. Astaxanthin can be detected as described above, and has maximal absorbance at 480 nm in acetone.

Yields of astaxanthin and other carotenoids can be increased by expression of a multifunctional geranylgeranyl pyrophosphate synthase, such as that from *S. shibatae* (SEQ ID NO:45) or an Archaeobacterial gene from *Archaeoglobus fulgidus* (GenBank Accession No. AF120272), in the engineered microorganism. The archaeobacteria GGPPS gene is a homolog of the endogenous *Rhodobacter* gene and encodes an enzyme that directly converts 3 IPP molecules and 1 DMAPP molecule to 1 GGPPS molecule, thereby reducing branching of the carotenoid pathway and eliminating production of other less desirable isoprenoids. Further reductions in less desirable metabolites can be obtained by eliminating endogenous bacteriochlorophyll biosynthesis, which redirects flow into carotenoid biosynthesis. For example, the *bchO*, *bchD*, and *bchI* genes can be deleted and/or replaced with an Archaeobacterial GGPPS gene. Additional increases in yield can be obtained by deletion of the endogenous *crtE* gene or the endogenous *crtC*, *crtD*, *crtE*, *crtA*, *crtI*, and *crtF* genes.

Common mutagenesis or knock-out technology can be used to delete endogenous genes. Alternatively, antisense technology can be used to reduce enzymatic activity. For example, a *R. sphaeroides* cell can be engineered to contain a cDNA that encodes an antisense molecule that prevents an enzyme from being made. The term "antisense molecule" as used herein encompasses any nucleic acid that contains sequences that correspond to the coding strand of an endogenous polypeptide. An antisense molecule also can have flanking sequences (e.g., regulatory sequences). Thus, antisense molecules



can be ribozymes or antisense oligonucleotides. A ribozyme can have any general structure including, without limitation, hairpin, hammerhead, or axhead structures, provided the molecule cleaves RNA.

5     *Control of the Ratio of Carotenoids*

      The amount of particular carotenoids, such as astaxanthin to canthaxanthin, or astaxanthin to zeaxanthin, can be controlled by expression of carotenoid genes from an inducible promoter or by use of constitutive promoters of different strengths. As used herein, "inducible" refers to both up-regulation and down regulation. An inducible  
10     promoter is a promoter that is capable of directly or indirectly activating transcription of one or more DNA sequences or genes in response to an inducer. In the absence of an inducer, the DNA sequences or genes will not be transcribed. The inducer can be a chemical agent such as a protein, metabolite, growth regulator, phenolic compound, or a physiological stress imposed directly by heat, cold, salt, or toxic elements, or indirectly  
15     through the action of a pathogen or disease agent such as a virus. The inducer also can be an illumination agent such as light, darkness and light's various aspects, which include wavelength, intensity, fluorescence, direction, and duration. Examples of inducible promoters include the lac system and the tetracycline resistance system from *E. coli*. In one version of the lac system, expression of lac operator-linked sequences is  
20     constitutively activated by a lacR-VP16 fusion protein and is turned off in the presence of IPTG. In another version of the lac system, a lacR-VP16 variant is used that binds to lac operators in the presence of IPTG, which can be enhanced by increasing the temperature of the cells.

      Components of the tetracycline (Tc) resistance system also can be used to regulate  
25     gene expression. For example, the Tet repressor (TetR), which binds to tet operator sequences in the absence of tetracycline and represses gene transcription, can be used to repress transcription from a promoter containing tet operator sequences. TetR also can be fused to the activation domain of VP 16 to create a tetracycline-controlled transcriptional activator (tTA), which is regulated by tetracycline in the same manner as TetR, i.e., tTA  
30     binds to tet operator sequences in the absence of tetracycline but not in the presence of

tetracycline. Thus, in this system, in the continuous presence of Tc, gene expression is repressed, and to induce transcription, Tc is removed.

Alternative methods of controlling the ratio of carotenoids include using enzyme inhibitors to regulate the activity levels of particular enzymes.

5

#### *Production of Carotenoids*

Carotenoids can be produced *in vitro* or *in vivo*. For example, one or more polypeptides of the invention can be contacted with an appropriate substrate or combination of substrates to produce the desired carotenoid (e.g., astaxanthin). See, FIG. 1 for a schematic of the carotenoid biosynthetic pathway.

10

A particular carotenoid (e.g., astaxanthin, lycopene,  $\beta$ -carotene, lutein, zeaxanthin, zeaxanthin diglucoside, or canthaxanthin) also can be produced by providing an engineered microorganism and culturing the provided microorganism with culture medium such that the carotenoid is produced. In general, the culture media and/or culture conditions are such that the microorganisms grow to an adequate density and produce the desired compound efficiently. For large-scale production processes, the following methods can be used. First, a large tank (e.g., a 100 gallon, 200 gallon, 500 gallon, or more tank) containing appropriate culture medium with, for example, a glucose carbon source is inoculated with a particular microorganism. After inoculation, the microorganisms are incubated to allow biomass to be produced. Once a desired biomass is reached, the broth containing the microorganisms can be transferred to a second tank. This second tank can be any size. For example, the second tank can be larger, smaller, or the same size as the first tank. Typically, the second tank is larger than the first such that additional culture medium can be added to the broth from the first tank. In addition, the culture medium within this second tank can be the same as, or different from, that used in the first tank. For example, the first tank can contain medium with xylose, while the second tank contains medium with glucose.

15

20

25

Once transferred, the microorganisms can be incubated to allow for the production of the desired carotenoid. Once produced, any method can be used to isolate the desired compound. For example, if the microorganism releases the desired carotenoid into the broth, then common separation techniques can be used to remove the biomass

30

from the broth, and common isolation procedures (e.g., extraction, distillation, and ion-exchange procedures) can be used to obtain the carotenoid from the microorganism-free broth. In addition, the desired carotenoid can be isolated while it is being produced, or it can be isolated from the broth after the product production phase has been terminated. If  
5 the microorganism retains the desired carotenoid, the biomass can be collected and the carotenoid can be released by treating the biomass or the carotenoid can be extracted directly from the biomass. Extracted carotenoid can be formulated as a nutraceutical. As used herein, a nutraceutical refers to a compound(s) that can be incorporated into a food, tablet, powder, or other medicinal form that, upon ingestion by a subject, provides a  
10 specific medical or physiological benefit to the subject.

Alternatively, the biomass can be collected and dried, without extracting the carotenoids. The biomass then can be formulated for human consumption (e.g., as a dietary supplement) or as an animal feed (e.g., for companion animals such as dogs, cats, and horses, or for production animals). For example, the biomass can be formulated for  
15 consumption by poultry such as chickens and turkeys, or by cattle, pigs, and sheep. Feeding of such compositions may increase yield of breast meat in poultry and may increase weight gain in other farm animals. In addition, the carotenoids may increase shelf-life of meat products due to the increased antioxidant protection afforded by the carotenoids. The biomass also can be formulated for use in aquaculture. For example,  
20 biomass that includes an engineered microorganism that is producing, e.g., astaxanthin and/or canthaxanthin, can be fed to fish or crustaceans to pigment the flesh or carapace, respectively. Such a composition is particularly useful for feeding to fish such as salmon, trout, sea bream, or snapper, or crustaceans such as shrimp, lobster, and crab.

One or more components can be added to the biomass before or after drying,  
25 including vitamins, other carotenoids, antioxidants such as ethoxyquin, vitamin E, butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT), or ascorbyl palmitate, vegetable oils such as corn oil, safflower oil, sunflower oil, or soybean oil, and an edible emulsifier, such as soy bean lecithin or sorbitan esters. Addition of antioxidants and vegetable oils can help prevent degradation of the carotenoid during processing (e.g.,  
30 drying), shipment, and storage of the composition.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

## EXAMPLES

### 5      Example 1 - Cloning of the zeaxanthin gene cluster from *Pantoea stewartii*:

Genomic DNA from *P. stewartii* was isolated and digested with restriction enzymes to yield genomic DNA fragments approximately 8-10 kB in size. These genomic DNA fragments were ligated into a vector cut with the same restriction enzyme, and electroporated into electrocompetent *E. coli*. Transformant colonies were individually  
 10      picked and transferred onto fresh solid media with the appropriate antibiotic selection (ampicillin/ampicillin substitute). It was thought that *E. coli* colonies containing the *P. stewartii* carotenoid genes would appear yellow in color due to the production of zeaxanthin pigment or red due to the production of lycopene. Although at least 2000 ampicillin resistant *E. coli* transformants were screened, none of the colonies were found  
 15      to contain the *P. stewartii* carotenoid genes.

Instead, a second, PCR based method was used to identify and sequence the carotenoid (crt) gene cluster from *P. stewartii* genomic DNA. Degenerate primers were designed based on homologous regions identified in the crt genes from *Erwinia herbicola* and *Erwinia uredovora*. Table 2 provides the position of the crt genes in *E. herbicola* and  
 20      *E. uredovora*.

**TABLE 2**  
**Position of crt genes in *E. herbicola* and *E. uredovora***

Gene name	Start of Gene (nucleotide #)		End of Gene (nucleotide #)	
	<i>E. herbicola</i>	<i>E. uredovora</i>	<i>E. herbicola</i>	<i>E. uredovora</i>
CrtE	3535	198	4458	1133
Orf-6	4521		5564	
CrtX	5561	1143	6802	2438
CrtY	6799	2422	7959	3570
CrtI	7956	3582	9434	5060
CrtB	9431	5096	10360	5986
CrtZ	10826 (complement)	6452 (complement)	10296 complement	5925 (complement)
Orf-12	12127 complement		10916 complement	

The following primers were designed (Table 3) and used in various combinations to yield PCR products of varying lengths. *P. stewartii* genomic DNA was used as template.

**TABLE 3**  
**Sequences of Degenerate Primers**

Primer Name	Primer Sequence	SEQ ID NO
P.s.BCHy1	5'-ATYATGCACGGCTGGGGWTGGSGMTGGCA - 3'	13
P.s. BCHy2	5' - GGCCARCGYTGATGCACCAGMCCGTCRTGCA - 3'	14
P.s.PS1	5' - CTGATGCTCTAYGCCTGGTGCCGCCA - 3'	15
P.s.PS2	5' - TCGCGRGCRATRTTSGTCARCTG - 3'	16
P.s.LBC1	5' - ATBMTSATGGAYGCSACSGT - 3'	17
P.s.LBC2	5' - YTRATCGARGAYACGCRCTA - 3'	18
P.s.LBC3	5' - RSGGCAGYGAATAGCCRG TG - 3'	19
P.s.LBC4	5' - AACAGCATSCGRTTCAGCAKGC GSA - 3'	20
P.s.PD5	5' - CCGACGGTKATCACCGATCC - 3'	21
P.s.PD6	5' - CTGCGCCSACCAGGTAGAG - 3'	22
P.s.GGPPS1	5' - CTYGACGAYATGCCCTGCATGGAC - 3' (MD92)	23
P.s.GGPPS2	5' - GTCGATTTWCCSGCGTCCTKATTG - 3' (MD93)	24

PCR was performed in a Gradient Thermocycler, and was started by incubating at 96°C for 5 minutes, followed by 40 cycles of denaturation at 96°C for 30 seconds, annealing at 40°C/45°C/50°C/55°C/or 60°C for 105 seconds, and extension at 72°C for 90 seconds, followed by incubation at 72°C for 10 mins. The concentration of MgCl<sub>2</sub> in the PCR reactions also was varied and ranged from a final concentration of 1.5 mM to 6 mM. Table 4 provides the predicted size of the PCR products with various primer combinations.

**TABLE 4**  
**Expected sizes of PCR Products**

Primer Combination	PCR product length (bp)	Product Observed
BCHy1/BCHy2	230	Yes
PS1/PS1	410	Yes
LBC1/LBC3	320	Yes
LBC1/LBC4	460	Yes
PD1/PD2	420	No
PD1/PD4	1260	No
LBC2/LBC3	240	No
PD3/PD4	410	Yes
LBC2/LBC4	380	Yes
PD5/PD6	1200	Yes
PS1/PS2	410	Yes
BCHy1/BCHy2	230	Yes
PsGGPPS1/PsGGPPS2	470	Yes
LBCDown1/PDUp1	470	Yes
PDDown1/PSUp1	300	Yes
BCHyDown1/PSDown1	700	Yes
LBCUp1/GGPPSdn1	1600	Yes

5 PCR reactions were electrophoresed through agarose gels to estimate sizes of PCR products and DNA was extracted from the gel using a Qiagen gel extraction kit. The purified PCR products were submitted to the Advanced Genetic Analysis Center (AGAC) at the University of Minnesota for sequencing. The obtained DNA sequences were subjected to BLAST analysis to determine if the sequences were homologous to *crt* genes from other bacteria. Sequence analysis of the 1.2-kb DNA fragment indicated that there was homology to phytoene desaturase (*crtI*) genes from *E. herbicola* and *E. uredovora*, while the 0.47 kb product had homology with the *crtE* genes from *E. herbicola* and *E. uredovora*.

15 Based on the DNA sequence information generated using the degenerate primers and amplified regions of the carotenoid genes from *P. stewartii*, primers specific for the *P. stewartii crt* genes were designed and are shown in Table 5. These specific primers were used to obtain information upstream and downstream of the DNA regions amplified

with the degenerate primers. This rationale was used to extend and obtain DNA sequence information about the *P. stewartii crt* genes.

**TABLE 5**  
***P. stewartii* primers**

Primer	Sequence	SEQ ID NO
PsOp.crtE	5'-GGCCGAATTCCAACGATGCTCTGGCAGTTA-3'	25
PSOp.crtZ(-)	5'-GGCCAGATCTACTTCAGGCGACGCTGAGAG-3'	26
PsOp.crtZ(+)	5'-GGCCAGATCTTACGCGCGGGTAAAGCCAAT-3'	27
PsOp.crtZ(2+)	5'-GGCCTCTAGAATTACCGCGTGGTTCTGAAG-3'	28
PsOp.crtZ(2-)	5'-GGCCTCTAGATCTGTACGCGCCACCGTTAT-3'	29

After unsuccessful attempts at completing the sequence *crt* gene cluster sequence from *P. stewartii* using PCR, the Universal Genome Walker kit from Clontech was used to obtain the complete the sequence of the *P. stewartii crtE* and *crtZ* genes. This kit uses a PCR based approach. The following primer pairs were synthesized and used for the genome walking experiments: GWcrtE2, 5' - CATCGGTAAGATCGTCAAGCAACTGAA - 3' (SEQ ID NO:30) and GWcrtE1, 5' - GATTTACCTGCATCCTGATTGATGTCT - 3' (SEQ ID NO:31); and GWcrtZ1, 5' - ATGTATAACCGTTTCAGGTAGCCTTTG - 3' (SEQ ID NO:32) and GWcrtZ2, 5' - AATACAGTAAACCATAAGCGGTCATGC - 3' (SEQ ID NO:33). The sequences of the *crt* genes and encoded proteins from *P. stewartii* were compared to the sequence of the *crt* genes and proteins from *E. herbicola* and *E. uredovora* using BLAST under default parameters. See, SEQ ID NOS 1-12 for the nucleotide and amino acid sequences of the *P. stewartii crt* genes. The results of the alignment are provided in Table 6.

**TABLE 6**  
**Comparison of crt genes and proteins from *P. stewartii* to *E. herbicola* and *E. uredovora***

Gene	Comparison of nucleotide sequence of <i>P. stewartii</i> to		Comparison of protein sequence of <i>P. stewartii</i> to	
	<i>E. herbicola</i>	<i>E. uredovora</i>	<i>E. herbicola</i>	<i>E. uredovora</i>
crtE	59%	80%	81%	83%
crtX	56%	75%	75%	74%
crtY	58%	77%	83%	82%

Gene	Comparison of nucleotide sequence of <i>P. stewartii</i> to		Comparison of protein sequence of <i>P. stewartii</i> to	
	<i>E. herbicola</i>	<i>E. uredovora</i>	<i>E. herbicola</i>	<i>E. uredovora</i>
crtI	69%	81%	89%	89%
crtB	63%	81%	88%	88%
crtZ	65%	84%	65%	88%

### **Example 2 - Cloning of a $\beta$ -carotene C4 Oxygenase from *Brevundimonas***

**aurantiaca**: Degenerate PCR primers for *crtW* were designed based on *crtW* genes from *Bradyrhizobium*, *Alcaligenes*, *Agrobacterium aurantiacum*, and *Paracoccus marcusii*.

- 5 The primers had the following sequences: (*crtW*(181P.m.) -  
 5'TTCATCATCGCGCATGAC3' (SEQ ID NO:34) and *crtW*(668P.m.)-  
 5'AGRTGRTGYTCGTGRTGA (SEQ ID NO:35), and were synthesized by Integrated  
 DNA Technologies Inc. (Coralville, IA). PCR was performed in a mastercycler gradient  
 machine (Eppendorf) with genomic DNA from *B. aurantiaca* (ATCC Accession No.  
 10 15266). Reaction conditions included five minutes at 96°C, followed by 30 cycles of  
 denaturation at 94°C for 30 sec., annealing at 50°C for 2 min., and extension at 72°C for 2  
 min 30 sec, and a final 72°C incubation for 10 min. An approximately 500-bp PCR  
 product was obtained and cloned into the vector pCR-BluntII-TOPO (Invitrogen Corp.  
 Carlsbad, CA).

- 15 Independent clones were sequenced using the universal M13 forward and reverse  
 primers. DNA sequencing was carried out at AGAC, University of Minnesota, St. Paul,  
 MN. Partial nucleotide sequence of the *crtW* gene was obtained. Alignment of the partial  
 sequence with known *crtW* genes indicated that the sequences aligned toward the  
 N-terminus and C-terminus, respectively, of the *crtW* genes from *Bradyrhizobium*,  
 20 *Alcaligenes*, *Agrobacterium aurantiacum*, and *Paracoccus marcusii*. The Universal  
 Genome Walker kit from Clontech was used to obtain the complete the sequence of the *B.*  
*aurantiaca crtW* gene. Primers were synthesized based on the partial sequence and used  
 for the genome walking experiments.

- 25 Upon obtaining sequence from the ends of the gene, the following oligonucleotide  
 primers were synthesized and used to amplify the complete *crtW* gene from genomic  
 DNA: 5'-GCGGCATAGGCTAGATTGAAG-3' (primer 1, Tm = 72°C, SEQ ID NO:36)  
 and 5'-GCGAGTTCCTTCTCACCTAT-3' (primer 2, Tm = 67°C, SEQ ID NO:37). *B.*



*aurantiaca* (ATCC 15266) genomic DNA was prepared with the Qiagen genomic-tip 500G kit (Valencia, CA; Catalog # 10262) following the manufacturers protocol. Briefly, 30 ml of *B. aurantiaca* culture were grown overnight at 30°C in ATCC medium 36 (Caulobacter medium; 2g/l peptone, 1 g/l yeast extract, 0.2 g/l MgSO<sub>4</sub>·7H<sub>2</sub>O). Cultures were harvested by centrifugation (15,000 x g; 10 minutes) and genomic DNA purified following the manufacturer's recommended protocol (Qiagen Genomic DNA Handbook for Blood, Cultured Cells, Tissue, Mouse Tails, Yeast, Bacteria (Gram- & some Gram+)). The Expand DNA polymerase system (Roche Molecular Biochemicals, Indianapolis, IN; catalog # 1732641) was used in a reaction that included 2 µl of *B. aurantiaca* genomic DNA (50 ng/µl), 1 µl of primer 1 (100 pmol/µl), 1 µl of primer 2 (100 pmol/µl), 5 µl of 10x PCR buffer, 1 µl of Expand DNA polymerase (3.5 U/µl), 2.5 µl of dimethyl sulfoxide (DMSO), 2 µl of dNTP's (10 nmol/µl each), and 35.5 µl of dd H<sub>2</sub>O. Reaction conditions included five minutes at 96°C, followed by 30 cycles of denaturation at 94°C for 30 sec., annealing at 50°C for 2 min., and extension at 72°C for 2 min 30 sec, and a final 72°C incubation for 10 min.

PCR products were electrophoresed through a 0.8% agarose gel and the ~0.85 kB band was excised from the gel and purified using the Qiagen QIAquick Gel Extraction Kit (catalog #28704) following the manufacturer's recommended protocol (QIAquick Spin Handbook). Gel-purified PCR product was cloned into the blunt-end cloning site of pCR-Blunt II-TOPO (Clontech; Palo Alto, CA) to generate pTOPOcrtW. Ligation mixtures were electroporated (25 µF, 200 Ohms, 12.5 KV/cm) into *E. coli* DH10B electromax cells (Gibco BRL; Gaithersburg, MD; catalog #18290-015). Transformants were allowed to recover 60 minutes at 37°C with shaking in 1 ml of SOC medium. Cells were plated on LB agar + 50 µg/ml kanamycin and allowed to grow overnight at 37°C. Transformant colonies were inoculated into 1 ml LB broth + 50 µg/ml kanamycin and allowed to grow overnight at 37°C with shaking. Minipreps were prepared using the QIAprep Spin Miniprep Kit (50) (catalog #27104) following the manufacturer's protocol and the presence of pTOPOcrtW was screened for by restriction analysis with *Eco*RI. *Eco*RI digests of pTOPOcrtW yielded products of ~0.85 Kbp and 3.5 Kbp.

The *crtW* gene was sequenced by AGAC, University of Minnesota, St. Paul, MN. The nucleotide sequence of the *crtW* gene from *B. aurantiaca* is provided in SEQ ID NO:38, and the protein encoded by the *crtW* gene is provided in SEQ ID NO:39.

5           **Example 3 - Transformation of pTOPOcrtW into *Pantoea stewartii* and production of astaxanthin and adonixanthin in *P.stewartii*::pTOPOcrtW:** The following protocol describes expression of *crtW* in the zeaxanthin producing host *P. stewartii*. This yields a transformed host that is capable of producing astaxanthin (i.e., 3,3'-dihydroxy- $\beta$ , $\beta$ -carotene-4,4'-dione) and adonixanthin (3,3'-dihydroxy- $\beta$ , $\beta$ -carotene-  
10 4-one). Electrocompetent *P. stewartii* (ATCC 8200) cells were prepared by culturing 50 ml of a 5% inoculum of *P. stewartii* cells in LB at 30°C -with agitation (250 rpm) until an OD<sub>590</sub> of 0.5-1.0 was reached. The bacteria were washed in 50 ml of 10mM HEPES (pH 7.0) and centrifuged for 10 minutes at 10,000xg. The wash was repeated with 25 ml of 10mM HEPES (pH 7.0) followed by the same centrifugation protocol. The cells then  
15 were washed once in 25 ml of 10% glycerol. Following centrifugation, the cells were resuspended in 500  $\mu$ l of 10% glycerol. Forty  $\mu$ l aliquots were frozen and kept at -80°C until use.

Plasmid TOPOcrtW was electroporated into electrocompetent *P. stewartii* cells (25  $\mu$ F, 25 KV/cm, 200 Ohms) and plated onto LB agar plates containing 50  $\mu$ g/ml  
20 kanamycin. As a negative control, pCR-Blunt II-TOPO self-ligated parental vector also was electroporated into *P. stewartii* and plated onto LB agar plates containing 50  $\mu$ g/ml kanamycin. Individual colonies of *P. stewartii*::pTOPOcrtW were screened by visual inspection for a phenotypic change from bright yellow pigmentation (production of zeaxanthin) to a reddish-orange pigmentation (production of astaxanthin) and chosen for  
25 further pigment analysis. No phenotypic change was noted for individual colonies of *P. stewartii*:: pCR-Blunt II-TOPO, so clones were randomly chosen for pigment analysis.

Production of astaxanthin was confirmed by HPLC/MS. Carotenoids were extracted from cells harvested from 5 day old cultures of *P. stewartii*::pTOPOcrtW or *P. stewartii*:: pCR-Blunt II-TOPO (25 ml) grown in LB with 50  $\mu$ g/ml kanamycin by  
30 resuspending the washed cell pellet in 5 ml of acetone. Glass beads were added and the mixture was incubated for 60 minutes at room temperature in the dark with occasional

vortexing. The cells were separated from the acetone extract by centrifugation at 15,000 x g for 10 minutes. The acetone supernatant then was analyzed by HPLC/MS.

A Waters 2790 LC system was used with two reverse-phase C30 specialty columns designed for carotenoid separation (YMCa Carotenoid S3m; 2.0 X 150 mm, 3 mm particle size; Waters Corporation, PN CT99S031502WT)), in tandem. The columns were run at room temperature. A gradient of Mobile Phase A (0.1% acetic acid) and Mobile Phase B (90% acetone) was used to separate zeaxanthin and astaxanthin according to the following gradient timetable: 0 min (10%A, 90%B), 10 min (100%B), 12 min (10%A, 90%B), 15 min (10%A, 90%B). Flow rate was 0.3 ml/min. Samples were stored at 20°C in an autosampler and a volume of 25 µL was injected. A Waters 996 Photodiode array detector, 350-550 nm, was used to detect zeaxanthin and astaxanthin. Under these chromatography conditions astaxanthin eluted at approximately 5.42-5.51 min and zeaxanthin eluted at approximately 6.22-6.4 min.

Carotenoid standards were used to identify the peaks. Astaxanthin was obtained from Sigma Chemical Co. (St. Louis, MO) and zeaxanthin was obtained from Extrasynthese (France). UV-Vis absorbtion spectra were used as diagnostic features for the carotenoids as were the molecular ion and fragmentation patterns generated using mass spectrometry. A positive-ion atmospheric pressure chemical ionization mass spectrometer was used; scan range, 400-800 m/z with a quadripole ion trap. A representative HPLC chromatogram is shown in FIG 3, which confirms production of astaxanthin in *P. stewartii* transformed with the *B. aurantiaca crtW* gene.

**Example 4 - Simultaneous Production of CoQ-10 and (3S, 3'S) Astaxanthin in a Microorganism:** Although *Phaffia rhodozyma* is not capable of producing the 3S, 3'S isoform of astaxanthin, it is known to produce Coenzyme Q-10. This compound has been found to have particularly high value as a nutraceutical. The current invention is of particular value since *R. sphaeroides* is known to produce Coenzyme Q-10 and has been transformed with genes that, while novel, are nevertheless homologous to native genes in the MABP. Consequently, the described organism can be expected to simultaneously produce both Coenzyme Q-10 and (3S, 3'S)-ATX. This is the first described production of the production of both (3S, 3'S)-ATX and Coenzyme Q-10 in a single microbial host.

The identification of (3S, 3'S)-ATX can be accomplished as described by Maoka, T., et al. J. Chromatogr. 318:122-124 (1985). Briefly, this consists of extraction of the carotenoid pigments by contacting the biomass with a suitable organic solvent such as acetone or dichloromethane. The carotenoid extract is then dried under a stream of liquid nitrogen and resuspended in a solvent of n-hexane-dichloromethane-ethanol (48:16:0.6). The extract is applied to a Sumipax OA-2000 (particle size 10 $\mu$ M) 250 x 4 mm I.D. (Sumitomo Chemicals, Osaka, Japan) chiral resolution HPLC column at a flow rate of 0.8 ml/min. Generally, the order of elution is expected to be (3R, 3'R)-ATX followed by (3R, 3'S; 3S, 3'R)-ATX followed by (3S, 3'S)-ATX. A similar separation is described in Maoka, T., et al. Comp. Biochem. Physiol. 83B:121-124 (1986). Briefly, this consists of isolation of the carotenoid, derivitization to the dibenzoate form with benzoyl chloride and separation of the enantiomers using a Sumipax OA-2000 chiral resolution HPLC column.

**Example 5 – Transformation of the multifunctional GGPP synthase from *Archeoglobus fulgidus* into *Rhodobacter* strain ppsr- with the *crtY* and *crtI* genes from *Pantoea stewartii* inserted into the chromosome:**

The following protocol describes the generation of a  $\beta$ -carotene producing strain of *R. sphaeroides* (ATCC 35053), a facultative photoheterotroph, in which the *ppsR* gene was deleted by using the in-frame deletion procedure of Higuchi, R., et al, Nucleic Acid Res. 16: 7351-7367 to generate strain  $\Delta$ REG. Table 7 describes the strains and plasmids used in this example. PpsR is a transcription factor that is involved in the repression of photosystem gene expression under aerobic growth conditions. The region of the chromosome that included the native *tspO*, *crtC*, *crtD*, *crtE* and *crtF* genes of  $\Delta$ REG were replaced by the lycopene  $\beta$  cyclase (*crtY*) and phytoene desaturase (*crtI*) genes from *P. stewartii* using the procedure of Oh and Kaplan, Biochemistry 38:2688-2696 (1999); and Lenz, et al., J. Bacteriology 176:4385-4393 (1994), to generate the strain  $\Delta$ REG( $\Delta$ 5:YI). Briefly, the *crtY* and *crtI* genes were cloned into pLO1, a suicide vector for *R. sphaeroides* containing the Kanamycin resistance gene and the *Bacillus subtilis* *sacB* gene encoding sensitivity to sucrose. DNA fragments flanking the *crtYI* genes and identical in sequence to ~500 bp internal fragments of the *R. sphaeroides* *tspO* and *crtF* genes were then cloned

into pLO1. These flanking DNA regions correspond to the desired region for insertion of the *crtYI* genes. Insertion of the *crtYI* genes in  $\Delta$ REG was confirmed using PCR analyses and appropriate PCR primers specific to the *crtYI* genes as well as flanking regions of the *R.sphaeroides* genome. The *crtYI* (*P. stewartii*) insertion and *tspO*, *crtC*, *crtD*, *crtE* and *crtF* (*R. sphaeroides*) deletion resulted in the lack of native carotenoid production and a change in the pigmentation from red to green, confirming the insertion event.

TABLE 7

Description of *Rhodobacter* Strains and Plasmids

Strain	Description	Major Carotenoid Produced	Comments
$\Delta$ REG	ATCC 35053; <i>ppsR</i> regulatory mutant	Sphaeroidenone (Native Carotenoid)	Regulatory mutant
$\Delta$ REG( $\Delta$ 5:YI)	<i>CrtY</i> and <i>crtI</i> genes of <i>P. stewartii</i> replaced 5 host genes ( <i>tspO</i> , <i>crtC</i> , <i>crtD</i> , <i>crtE</i> and <i>crtF</i> ) on chromosome	None	$\beta$ -carotene biosynthetic genes placed in chromosome. No carotenoid production because of <i>crtE</i> deletion
$\Delta$ REG( $\Delta$ 5:YI)::pPctrl	Control vector introduced into $\Delta$ REG( $\Delta$ 5:YI) host	None	Control vector contains <i>rrnB</i> promoter but no biosynthetic genes
$\Delta$ REG( $\Delta$ 5:YI)::pPgps	<i>gps</i> gene of <i>A. fulgidus</i> inserted into pPctrl control vector and introduced into $\Delta$ REG( $\Delta$ 5:YI) host	$\beta$ -Carotene	<i>gps</i> gene on plasmid complements <i>crtE</i> deletion. Complete pathway for $\beta$ -carotene production

Strain	Description	Major Carotenoid Produced	Comments
$\Delta$ REG( $\Delta$ 5:YI) ( $\Delta$ A:gps)	<i>gps</i> gene of <i>A. fulgidus</i> replaced <i>crtA</i> host gene on chromosome of $\Delta$ REG( $\Delta$ 5:YI) host	$\beta$ -Carotene	<i>gps</i> gene inserted into genome complements <i>crtE</i> deletion. Complete pathway for $\beta$ -carotene production
$\Delta$ REG( $\Delta$ 5:YI) ( $\Delta$ A:gps) ::pPWZ	<i>crtW</i> and <i>crtZ</i> genes inserted into pPctrl control vector and introduced into $\Delta$ REG( $\Delta$ 5:YI) ( $\Delta$ A:gps) host	Astaxanthin	<i>crtW</i> and <i>crtZ</i> genes convert $\beta$ -carotene into astaxanthin
$\Delta$ REG( $\Delta$ 5:YI) ( $\Delta$ A:gps) ::pPgpsWZ	<i>gps</i> , <i>crtW</i> and <i>crtZ</i> genes inserted into pPctrl control vector and introduced into $\Delta$ REG( $\Delta$ 5:YI) ( $\Delta$ A:gps) host	Astaxanthin	Additional copies of <i>A. fulgidus gps</i> gene on plasmid increases production of astaxanthin
Plasmids	Genetic elements inserted		
PBBR1MCS2	None		
PPctrl	<i>rrnB</i> promoter		
PPgps	<i>rrnB</i> promoter, <i>A. fulgidus gps</i>		
PPWZ	<i>rrnB</i> promoter, <i>P. stewartii crtZ</i> , <i>B. aurantiacum crtW</i>		
PPgpsWZ	<i>rrnB</i> promoter, <i>A. fulgidus gps</i> <i>P. stewartii crtZ</i> , <i>B. aurantiacum crtW</i>		

The pPctrl vector was constructed by inserting a copy of the *R. sphaeroides rrnB* promoter (GenBank Accession # X53854; rrnBP) into the vector pBBR1MCS2 (GenBank Accession # U23751). The *rrnB* promoter was isolated from the vector pTEX24 (S. Kaplan) by a *Bam*HI restriction enzyme digest, which released the promoter as a 363 bp fragment. This fragment was gel purified from a 2% Tris-acetate-EDTA (TAE) agarose gel. To prepare the pBBR1MCS2 vector for ligation, it also was digested with *Bam*HI

and the enzyme heat inactivated at 80°C for 20 minutes. The digested vector was dephosphorylated with shrimp alkaline phosphatase (Roche Molecular Biochemicals, Indianapolis, IN), and gel purified from a 1% TAE-agarose gel. The prepared vector and the *rrnB* fragment were ligated using T4 DNA ligase at 16°C for 16 hours to generate the plasmid pPctrl. One  $\mu$ L of ligation reaction was used to electroporate 40  $\mu$ L of *E. coli* Electromax™ DH10BTM cells (Life Technologies, Inc., Rockville, MD).

Electroporated cells were plated on LB media containing 25  $\mu$ g/mL of kanamycin (LBK). pPctrl DNA was isolated from cultures of single colonies and was digested with *Hind* III to confirm the presence of a single insertion of the *rrnB* promoter. The sequence of pPctrl also was confirmed by DNA sequencing.

The multifunctional GGPP synthase (*gps*) gene from *A. fulgidus* (GenBank Accession No. AF120272) was cloned into the multiple cloning site of pPctrl to generate the construct pPgps.

Electrocompetent  $\Delta$ REG( $\Delta$ 5:YI) cells were prepared as follows: 5 ml cultures were inoculated using Sistrom's media supplemented with trace elements, vitamins (O'Gara, et al., *J. Bacteriol.* 180:4044-4050 (1988); Cohen-Bazire, et al. *J. Cell. Comp. Physiol.* 49:25-68 (1957)) and 0.4% glucose as a carbon source, and grown overnight at 30°C with shaking. This culture was diluted 1/100 in 300 mL of the same media and grown to an OD<sub>660</sub> of 0.5-0.8. The cells were chilled on ice for 10 minutes and then centrifuged for 6 minutes at 7,500 g. The supernatant was discarded and the cell pellet was resuspended in ice-cold 10% glycerol at half of the original volume. The cells were pelleted by centrifugation for 6 minutes at 7,500 g. The supernatant was again discarded and cells were resuspended in ice cold 10% glycerol at one quarter of the original volume. The last centrifugation and resuspension steps were repeated, followed by centrifugation for 6 minutes at 7,500 g. The supernatant was decanted and the cells resuspended in the small volume of glycerol that did not drain out. Additional ice-cold 10% glycerol was added to resuspend the cells if necessary. Forty  $\mu$ L of the resuspended cells was used in a test electroporation (see below) to determine if the cells needed to be concentrated by centrifugation or diluted with 10% ice-cold glycerol. Time constants of 8.5-9.0 resulted in good transformation efficiencies. Once an acceptable time constant was achieved, cells

were aliquoted into cold microfuge tubes and stored at -80°C. All water used for media and glycerol was 18 Mohm or higher.

Electroporation of  $\Delta$ REG( $\Delta$ 5:YI) was carried out as follows. One  $\mu$ L of pPgps or pPctrl vector DNA was gently mixed into 40  $\mu$ L of  $\Delta$ REG( $\Delta$ 5:YI) electrocompetent cells, which then were transferred to an electroporation cuvette with a 0.2 cm electrode gap. Electroporations were conducted using a Biorad Gene Pulser II (Biorad, Hercules, CA) with settings at 2.5 kV of potential, 400 ohms of resistance, and 25  $\mu$ F of capacitance. Cells were recovered in 400  $\mu$ L SOC media at 30°C for 6-16 hours. The cells were then plated, 200  $\mu$ L per plate, on LB medium containing 50  $\mu$ g/ml kanamycin and incubated at 30°C for 5-6 days.

After incubation, greenish colonies were observed on plates of  $\Delta$ REG( $\Delta$ 5:YI) transformed with pPctrl plasmid DNA. The colonies that appeared on plates of  $\Delta$ REG( $\Delta$ 5:YI) transformed with pPgps plasmid DNA appeared yellow. The yellow pigmentation was indicative of  $\beta$ -carotene production in  $\Delta$ REG( $\Delta$ 5:YI) expressing the *A. fulgidus* *gps* gene from pPgps.

Single yellow colonies were grown up in Siström's liquid media supplemented with vitamins, trace elements and 0.4% glucose as well as 50  $\mu$ g/ml kanamycin, at 30°C with shaking for 24-48 hours. Carotenoids were extracted and subjected to LCMS analysis as described above. Under the chromatography conditions used,  $\beta$ -carotene eluted at approximately 13.87-14.2 min.  $\beta$ -carotene standard (Sigma chemical, St. Louis, MO) was used to identify the peaks. The UV-Vis absorption spectra and the retention time using HPLC were used as diagnostic features for  $\beta$ -carotene identification in  $\Delta$ REG( $\Delta$ 5:YI) transformed with pPgps DNA, as well as the molecular ion and fragmentation patterns generated during mass spectrometry. Thus, the production of  $\beta$ -carotene was confirmed in  $\Delta$ REG( $\Delta$ 5:YI) expressing the *A. fulgidus* *gps* gene from pPgps.

**Example 6 – Transformation of the  $\beta$ -carotene C-4 ketolase (*crtW*) gene from *Brevumdimonas aurantiacum* and  $\beta$ -carotene hydroxylase (*crtZ*) from *P. stewartii* into the  $\Delta$ REG( $\Delta$ 5:Y1) strain of *Rhodobacter* with the *gps* gene from *Archeoglobus fulgidus* inserted into the chromosome:** The following protocol describes the



generation of an astaxanthin producing strain of *R. sphaeroides* using  $\Delta$ REG( $\Delta$ 5:YI), described above. See also Table 7 for further description of the strains and plasmids that were used in this example. Using the gene insertion method described by Higuchi, R., et al, Nucleic Acid Res. 16: 7351-7367, the *crtA* gene of  $\Delta$ REG( $\Delta$ 5:YI) was replaced by the

5

*gps* gene from *A. fulgidus* to generate the strain  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps). Electrocompetent cells  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps) were generated as described above.

The construct pPggsWZ was produced by cloning the *crtW* gene from *B. aurantiacum*, the *crtZ* gene from *P.stewartii*, and the *gps* gene from *A fulgidus* into the pPctrl plasmid using appropriate restriction enzymes. The construct pPWZ was produced

10

by cloning the *crtW* gene from *B. aurantiacum* and the *crtZ* gene from *P.stewartii* into the pPctrl plasmid using appropriate restriction enzymes. The pPWZ or pPggsWZ constructs were electroporated into electrocompetent  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps) as described earlier to generate  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPWZ or  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPggsWZ, respectively. Transformation mixtures were plated

15

out onto LB plates containing 50  $\mu$ g/ml kanamycin. PCR analyses using PCR primers specific for *crtZ* were used to confirm the presence of the pPWZ or pPggsWZ plasmids in  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps). Single colonies of  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPWZ or  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPggsWZ were grown up in media supplemented with 50  $\mu$ g/ml kanamycin as described earlier. Cell pellets were washed with distilled water and then carotenoids were extracted using acetone:methanol (7:2) at 30°C for 30 mins with shaking at 225 rpm. Carotenoid analysis was performed using LCMS analysis described above. The UV-Vis absorption spectra and the retention time using HPLC were used as diagnostic features for astaxanthin identification in  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPWZ and

25

$\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPggsWZ, as well as the molecular ion and fragmentation patterns generated during mass spectrometry. The production of astaxanthin was confirmed in both  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPWZ and  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPggsWZ. Increased astaxanthin production was observed in  $\Delta$ REG( $\Delta$ 5:YI)( $\Delta$ A:gps)::pPggsWZ.

**Example 7: Cloning and sequencing of a novel multifunctional****Geranylgeranyl pyrophosphate synthase gene (*gps*) from *Sulfolobus shibatae*:**

Degenerate primer sequences MFGGPP1 (5'CCA YGAYGAYATWATGGA3', SEQ ID NO:40) and MFGGPP2 (5'YTTYTTVCCYTYCCTAAT3', SEQ ID NO:41) were  
5 designed based on conserved sequences in *gps* gene sequences from *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius* and synthesized by Integrated DNA Technologies (Coralville, IA). PCR was performed in a mastercycler gradient machine (Eppendorf) with genomic DNA from *S. shibatae* (ATCC Accession No. 51178, lot # 1162977). Reaction conditions included five minutes at 96°C, followed by 30 cycles of  
10 denaturation at 94°C for 30 sec., annealing at 50 + 10°C for 60 sec., and extension at 72°C for 90 sec., and a final 72°C incubation for 10 min. An approximately 500-bp PCR product was obtained and cloned into the vector pC-BuntII-TOPO (Invitrogen Corp. Carlsbad, CA).

Independent clones were sequenced using the universal M13 forward and reverse  
15 primers. DNA sequencing was carried out at the AGAC, University of Minnesota, St. Paul, MN. DNA sequence analysis of this PCR product indicated similarity to the *gps* genes from *S. solfataricus* and *S. acidocaldarius*. The Universal Genome Walker kit (Clontech) was used to obtain more of the *gps* gene sequence flanking the original PCR product from *S. shibatae*. Primers were synthesized based on the partial sequence and  
20 used for genome walking experiments.

The following strategy was used to completely sequence the *S. shibatae gps* gene. The ERWCRTS homolog was observed upstream of the *S. solfataricus gps* gene. The UDP-A-acetylglucosamine—Dolichyl-phosphate-N-acetylglucosamine  
phosphotransferase gene was present downstream of the *gps* gene in both *S. solfataricus*  
25 and *S. acidocaldarius*. Primers were designed based on the sequence of the two genes SsDolidn (5'ACAGCGTTGGACACTCAG 3', SEQ ID NO:42) and SsERCRTup (5' GCGTCGATAATGGAAGTGAG 3', SEQ ID NO:43) of the *gps* gene. An approximately 2 kb PCR product was amplified using the SsDolidn and SsERCRTup primers and genomic DNA from *S. shibatae*. This PCR product was cloned into the vector pC-BuntII-  
30 TOPO as described above and sequenced using the universal M13 forward and reverse primers. The nucleotide sequence of the *gps* gene from *S. shibatae* is presented in SEQ

ID NO: 44, and the amino acid sequence of the protein encoded by the *gps* gene is presented in SEQ ID NO:45.

#### OTHER EMBODIMENTS

- 5           It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

**WHAT IS CLAIMED IS:**

1. An isolated nucleic acid having at least 76% sequence identity to the nucleotide sequence of SEQ ID NO:1 or to a fragment of SEQ ID NO:1 at least 33 contiguous nucleotides in length.  
5
2. The isolated nucleic acid of claim 1, said nucleic acid having at least 80% sequence identity to the nucleotide sequence of SEQ ID NO:1.
3. The isolated nucleic acid of claim 1, said nucleic acid having at least 85% sequence identity to the nucleotide sequence of SEQ ID NO:1.  
10
4. The isolated nucleic acid of claim 1, said nucleic acid having at least 90% sequence identity to the nucleotide sequence of SEQ ID NO:1.
5. The isolated nucleic acid of claim 1, said nucleic acid having at least 95% sequence identity to the nucleotide sequence of SEQ ID NO:1.  
15
6. An expression vector comprising the nucleic acid of claim 1 operably linked to an expression control element.  
20
7. An isolated nucleic acid encoding a zeaxanthin glucosyl transferase polypeptide at least 75% identical to the amino acid sequence of SEQ ID NO:2.
8. An isolated nucleic acid having at least 78% sequence identity to the nucleotide sequence of SEQ ID NO:3 or to a fragment of SEQ ID NO:3 at least 32 contiguous nucleotides in length.  
25
9. The isolated nucleic acid of claim 8, said nucleic acid having at least 80% sequence identity to the nucleotide sequence of SEQ ID NO:3.  
30

10. The isolated nucleic acid of claim 8, said nucleic acid having at least 85% sequence identity to the nucleotide sequence of SEQ ID NO:3.
11. The isolated nucleic acid of claim 8, said nucleic acid having at least 90% sequence  
5 identity to the nucleotide sequence of SEQ ID NO:3.
12. The isolated nucleic acid of claim 8, said nucleic acid having at least 95% sequence identity to the nucleotide sequence of SEQ ID NO:3.
- 10 13. An expression vector comprising the nucleic acid of claim 8 operably linked to an expression control element.
14. An isolated nucleic acid encoding a lycopene  $\beta$ -cyclase polypeptide at least 83% identical to the amino acid sequence of SEQ ID NO:4.
- 15 15. An isolated nucleic acid having at least 81% sequence identity to the nucleotide sequence of SEQ ID NO:5 or to a fragment of SEQ ID NO:5 at least 60 contiguous nucleotides in length.
- 20 16. The isolated nucleic acid of claim 15, said nucleic acid having at least 85% sequence identity to the nucleotide sequence of SEQ ID NO:5.
17. The isolated nucleic acid of claim 15, said nucleic acid having at least 90% sequence identity to the nucleotide sequence of SEQ ID NO:5.
- 25 18. The isolated nucleic acid of claim 15, said nucleic acid having at least 95% sequence identity to the nucleotide sequence of SEQ ID NO:5.
19. An expression vector comprising the nucleic acid of claim 15 operably linked to an  
30 expression control element.

20. An isolated nucleic acid encoding a geranylgeranyl pyrophosphate synthase polypeptide at least 85% identical to the amino acid sequence of SEQ ID NO:6.
21. An isolated nucleic acid having at least 82% sequence identity to the nucleotide  
5 sequence of SEQ ID NO:7 or to a fragment of SEQ ID NO:7 at least 30 contiguous nucleotides in length.
22. The isolated nucleic acid of claim 21, said nucleic acid having at least 85% sequence identity to the nucleotide sequence of SEQ ID NO:7.
- 10 23. The isolated nucleic acid of claim 21, said nucleic acid having at least 90% sequence identity to the nucleotide sequence of SEQ ID NO:7.
24. The isolated nucleic acid of claim 21, said nucleic acid having at least 95% sequence  
15 identity to the nucleotide sequence of SEQ ID NO:7.
25. An expression vector comprising the nucleic acid of claim 21 operably linked to an expression control element.
- 20 26. An isolated nucleic acid encoding a phytoene desaturase polypeptide at least 90% identical to the amino acid sequence of SEQ ID NO:8.
27. An isolated nucleic acid having at least 82% sequence identity to the nucleotide  
25 sequence of SEQ ID NO:9 or to a fragment of SEQ ID NO:9 at least 23 contiguous nucleotides in length.
28. The isolated nucleic acid of claim 27, said nucleic acid having at least 85% sequence identity to the nucleotide sequence of SEQ ID NO:9.
- 30 29. The isolated nucleic acid of claim 27, said nucleic acid having at least 90% sequence identity to the nucleotide sequence of SEQ ID NO:9.

30. The isolated nucleic acid of claim 27, said nucleic acid having at least 95% sequence identity to the nucleotide sequence of SEQ ID NO:9.
- 5 31. An expression vector comprising the nucleic acid of claim 27 operably linked to an expression control element.
32. An isolated nucleic acid encoding a phytoene synthase polypeptide at least 89% identical to the amino acid sequence of SEQ ID NO:10.
- 10 33. An isolated nucleic acid having at least 85% sequence identity to the nucleotide sequence of SEQ ID NO:11 or to a fragment of SEQ ID NO:11 at least 36 contiguous nucleotides in length.
- 15 34. The isolated nucleic acid of claim 33, said nucleic acid having at least 85% sequence identity to the nucleotide sequence of SEQ ID NO:11.
35. The isolated nucleic acid of claim 33, said nucleic acid having at least 90% sequence identity to the nucleotide sequence of SEQ ID NO:11.
- 20 36. The isolated nucleic acid of claim 33, said nucleic acid having at least 95% sequence identity to the nucleotide sequence of SEQ ID NO:11.
37. An expression vector comprising the nucleic acid of claim 33 operably linked to an expression control element.
- 25 38. An isolated nucleic acid encoding a  $\beta$ -carotene hydroxylase polypeptide at least 90% identical to the amino acid sequence of SEQ ID NO:12.
- 30 39. Membranous bacteria comprising at least one exogenous nucleic acid encoding phytoene desaturase, lycopene  $\beta$ -cyclase,  $\beta$ -carotene hydroxylase, and  $\beta$ -carotene C4

oxygenase, wherein expression of said at least one exogenous nucleic acid produces detectable amounts of astaxanthin in said membranous bacteria.

- 5        40. The membranous bacteria of claim 39, wherein the amino acid sequence of said phytoene desaturase is at least 90% identical to the amino acid sequence of SEQ ID NO:8.
- 10       41. The membranous bacteria of claim 39, wherein the amino acid sequence of said lycopene  $\beta$ -cyclase is at least 83% identical to the amino acid sequence of SEQ ID NO:4.
- 15       42. The membranous bacteria of claim 39, wherein the amino acid sequence of said  $\beta$ -carotene hydroxylase is at least 90% identical to the amino acid sequence of SEQ ID NO:12.
- 20       43. The membranous bacteria of claim 39, wherein said membranous bacteria further comprises an exogenous nucleic acid encoding geranylgeranyl pyrophosphate synthase.
- 25       44. The membranous bacteria of claim 39, wherein said membranous bacteria lacks endogenous bacteriochlorophyll biosynthesis.
45. The membranous bacteria of claim 43, wherein said exogenous nucleic acid encodes a multifunctional geranylgeranyl pyrophosphate synthase.
46. The membranous bacteria of claim 45, wherein the amino acid sequence of said multifunctional geranylgeranyl pyrophosphate synthase is at least 90% identical to the amino acid sequence of SEQ ID NO:45.



47. The membranous bacteria of claim 39, wherein the amino acid sequence of said  $\beta$ -carotene C4 oxygenase is at least 80% identical to the amino acid sequence of SEQ ID NO:39.
- 5      48. The membranous bacteria of claim 39, wherein said membranous bacteria further comprise an exogenous nucleic acid encoding phytoene synthase.
49. The membranous bacteria of claim 48, wherein the amino acid sequence of said phytoene synthase is at least 89% identical to the amino acid sequence of SEQ ID  
10      NO:10.
50. The membranous bacteria of claim 39, wherein said membranous bacteria are a *Rhodobacter* species.
- 15      51. Membranous bacteria, said membranous bacteria comprising an exogenous nucleic acid encoding a phytoene desaturase having an amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:8, and wherein said membranous bacteria produces detectable amounts of lycopene.
- 20      52. The membranous bacteria of claim 51, wherein said membranous bacteria further comprise a lycopene  $\beta$ -cyclase, and wherein said membranous bacteria produce detectable amounts of  $\beta$ -carotene.
53. The membranous bacteria of claim 52, wherein said membranous bacteria further  
25      comprise a  $\beta$ -carotene hydroxylase, and wherein said membranous bacteria produce detectable amounts of zeaxanthin.
54. Membranous bacteria comprising at least one exogenous nucleic acid encoding phytoene desaturase, lycopene  $\beta$ -cyclase, and  $\beta$ -carotene C4 oxygenase, wherein  
30      expression of said at least one exogenous nucleic acid produces detectable amounts of canthaxanthin in said membranous bacteria.

55. A composition comprising an engineered *Rhodobacter* cell, wherein said cell produces a detectable amount of astaxanthin or canthaxanthin.
- 5 56. The composition of claim 55, wherein said engineered *Rhodobacter* cell comprises at least one exogenous nucleic acid encoding phytoene desaturase, lycopene  $\beta$ -cyclase,  $\beta$ -carotene hydroxylase, and  $\beta$ -carotene C4 oxygenase.
57. The composition of claim 55, wherein said composition is formulated for aquaculture.
- 10 58. The composition of claim 57, wherein said composition pigments the flesh of fish or the carapace of crustaceans after ingestion.
59. The composition of claim 55, wherein said composition is formulated for human consumption.
- 15 60. The composition of claim 55, wherein said composition is formulated as an animal feed.
- 20 61. The composition of claim 60, wherein said animal feed is formulated for consumption by chickens, turkeys, cattle, swine, or sheep.
62. A method of making a nutraceutical, said method comprising extracting carotenoids from an engineered *Rhodobacter* cell, said engineered *Rhodobacter* cell comprising at least one exogenous nucleic acid encoding phytoene desaturase, lycopene  $\beta$ -cyclase,  $\beta$ -carotene hydroxylase, and  $\beta$ -carotene C4 oxygenase, and wherein said *Rhodobacter* cell produces detectable amounts of astaxanthin.
- 25 63. Membranous bacteria, said membranous bacteria comprising an exogenous nucleic acid encoding a lycopene  $\beta$ -cyclase having an amino acid sequence at least 83% identical to the amino acid sequence of SEQ ID NO:4.
- 30

64. The membranous bacteria of claim 63, said membranous bacteria further comprising a phytoene desaturase, wherein said membranous bacteria produces detectable amounts of  $\beta$ -carotene.
- 5
65. The membranous bacteria of claim 64, said membranous bacteria further comprising a  $\beta$ -carotene hydroxylase, wherein said bacteria produces detectable amounts of zeaxanthin.
- 10
66. Membranous bacteria, said membranous bacteria comprising a  $\beta$ -carotene hydroxylase having an amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:12.
- 15
67. The membranous bacteria of claim 66, said membranous bacteria further comprising a lycopene  $\beta$ -cyclase, and wherein said membranous bacteria produces detectable amounts of zeaxanthin.
- 20
68. The membranous bacteria of claim 67, said membranous bacteria further comprising a phytoene desaturase, wherein said membranous bacteria produces detectable amounts of  $\beta$ -carotene.
- 25
69. Membranous bacteria, said bacteria lacking an endogenous nucleic acid encoding a farnesyl pyrophosphate synthase, and wherein said bacteria produce detectable amounts of carotenoids.
70. The membranous bacteria of claim 69, wherein said bacteria comprise an exogenous nucleic acid encoding a multifunctional geranylgeranyl pyrophosphate synthase.
- 30
71. The membranous bacteria of claim 70, wherein the amino acid sequence of said multifunctional geranylgeranyl pyrophosphate synthase is at least 90% identical to the amino acid sequence of SEQ ID NO:45.

72. The membranous bacteria of claim 69, wherein said membranous bacteria are a species of *Rhodobacter*.
- 5 73. An isolated nucleic acid having at least 60% sequence identity to the nucleotide sequences of SEQ ID NO:38, or to a fragment of the nucleic acid of SEQ ID NO:38 at least 15 contiguous nucleotides in length.
- 10 74. The isolated nucleic acid of claim 73, said nucleic acid having at least 80% sequence identity to the nucleotide sequences of SEQ ID NO:38, or to a fragment of the nucleic acid of SEQ ID NO:38 at least 15 contiguous nucleotides in length.
- 15 75. The isolated nucleic acid of claim 73, said nucleic acid having at least 90% sequence identity to the nucleotide sequences of SEQ ID NO:38, or to a fragment of the nucleic acid of SEQ ID NO:38 at least 15 contiguous nucleotides in length.
76. The isolated nucleic acid of claim 73, wherein said nucleic acid encodes a  $\beta$ -carotene C4 oxygenase.
- 20 77. Membranous bacteria comprising an exogenous nucleic acid encoding a  $\beta$ -carotene C4 oxygenase, said  $\beta$ -carotene oxygenase having an amino acid sequence at least 80% identical to the amino acid sequence of SEQ ID NO:39.
- 25 78. A host cell comprising an exogenous nucleic acid, wherein the exogenous nucleic acid comprises a nucleic acid sequence encoding one or more polypeptides that catalyze the formation of (3S, 3'S) astaxanthin, wherein the host cell produces CoQ-10 and (3S, 3'S) astaxanthin.
- 30 79. A method of making CoQ-10 and (3S, 3'S) astaxanthin at substantially the same time, the method comprising transforming a host cell with a nucleic acid, wherein the nucleic acid comprises a nucleic acid sequence that encodes one or more

polypeptides, wherein the polypeptides catalyze the formation of (3S, 3'S) astaxanthin; and culturing the host cell under conditions that allow for the production of (3S, 3'S) astaxanthin and CoQ-10.

- 5      80. The method of claim 79, additionally comprising transforming the host cell with at least one exogenous nucleic acid, the exogenous nucleic acid encoding one or more polypeptides, wherein the polypeptides catalyze the formation of CoQ-10.
- 10      81. An isolated nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:38, and SEQ ID NO:44.
- 15      82. An isolated nucleic acid having at least 90% sequence identity to the nucleotide sequences of SEQ ID NO:44, or to a fragment of the nucleic acid of SEQ ID NO:44 at least 60 contiguous nucleotides in length.
- 20      83. A method of making geranylgeranyl pyrophosphate, said method comprising contacting isopentenyl pyrophosphate and dimethylallyl pyrophosphate with a polypeptide encoded by the isolated nucleic acid of claim 82.
- 25      84. A method of making geranylgeranyl pyrophosphate, said method comprising contacting farnesyl pyrophosphate and isopentenyl pyrophosphate with a polypeptide encoded by the isolated nucleic acid of claim 15 or the polypeptide of claim 20.
- 30      85. A method of making  $\beta$ -carotene, said method comprising contacting lycopene with a polypeptide encoded by the isolated nucleic acid of claim 8 or the polypeptide of claim 14.
- 30      86. A method of making lycopene, said method comprising contacting phytoene with a polypeptide encoded by the isolated nucleic acid of claim 21 or the polypeptide of claim 26.

87. A method of making phytoene, said method comprising contacting geranylgeranyl pyrophosphate with a polypeptide encoded by the isolated nucleic acid of claim 27 or the polypeptide of claim 32.
- 5
88. A method of making zeaxanthin, said method comprising contacting  $\beta$ -carotene with a polypeptide encoded by the isolated nucleic acid of claim 33 or the polypeptide of claim 38.
- 10
89. A method of making canthaxanthin, said method comprising contacting  $\beta$ -carotene with a polypeptide encoded by the isolated nucleic acid of claim 73 or a polypeptide having an amino acid sequence at least 80% identical to the amino acid sequence of SEQ ID NO:39.
- 15
90. A method of making astaxanthin, said method comprising contacting canthaxanthin with a polypeptide encoded by the isolated nucleic acid sequence of claim 33 or the polypeptide of claim 38.
- 20
91. A method of making astaxanthin, said method comprising contacting zeaxanthin with a polypeptide encoded by the isolated nucleic acid sequence of claim 73 or a polypeptide having an amino acid sequence at least 80% identical to the amino acid sequence of SEQ ID NO:39.

Figure 1

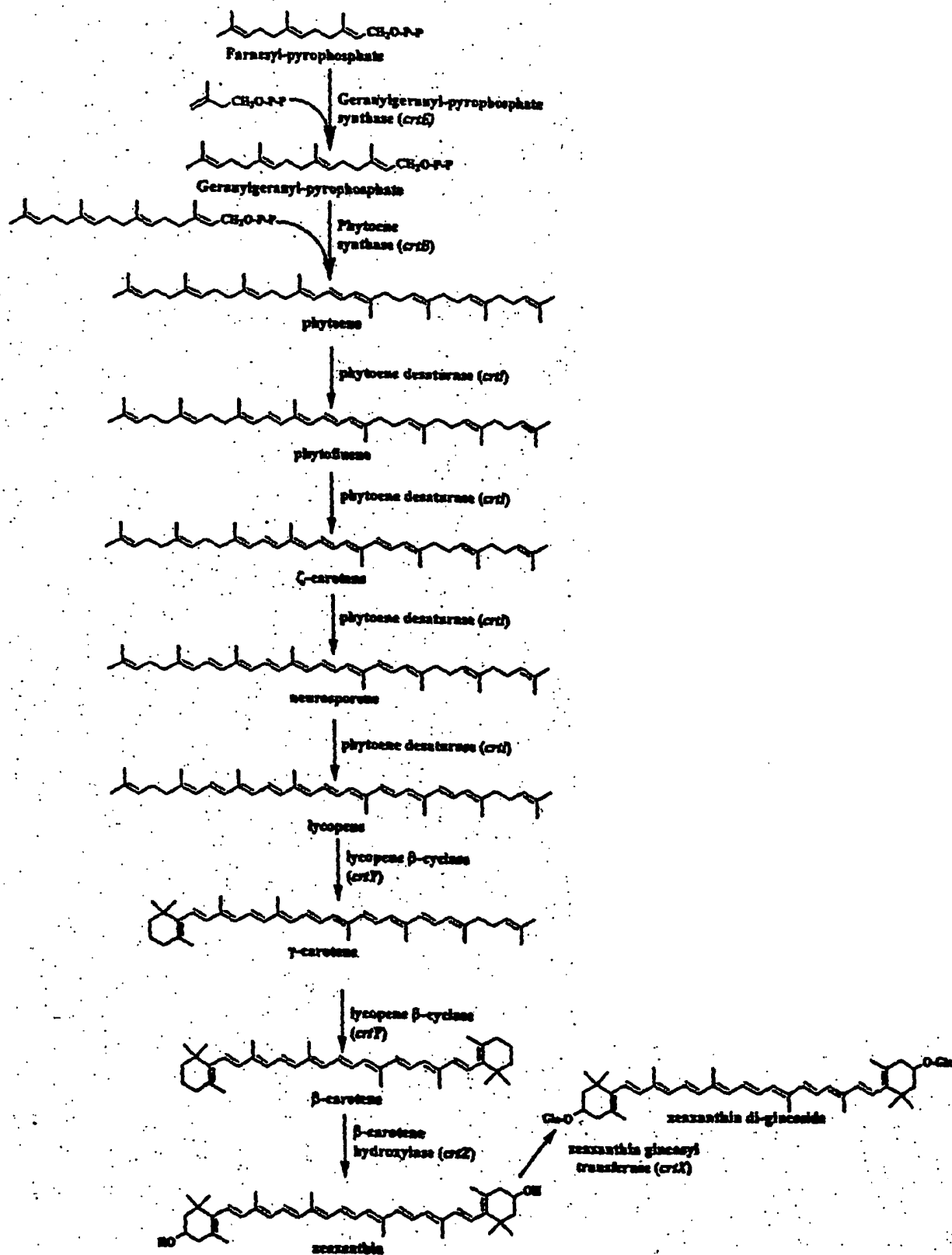
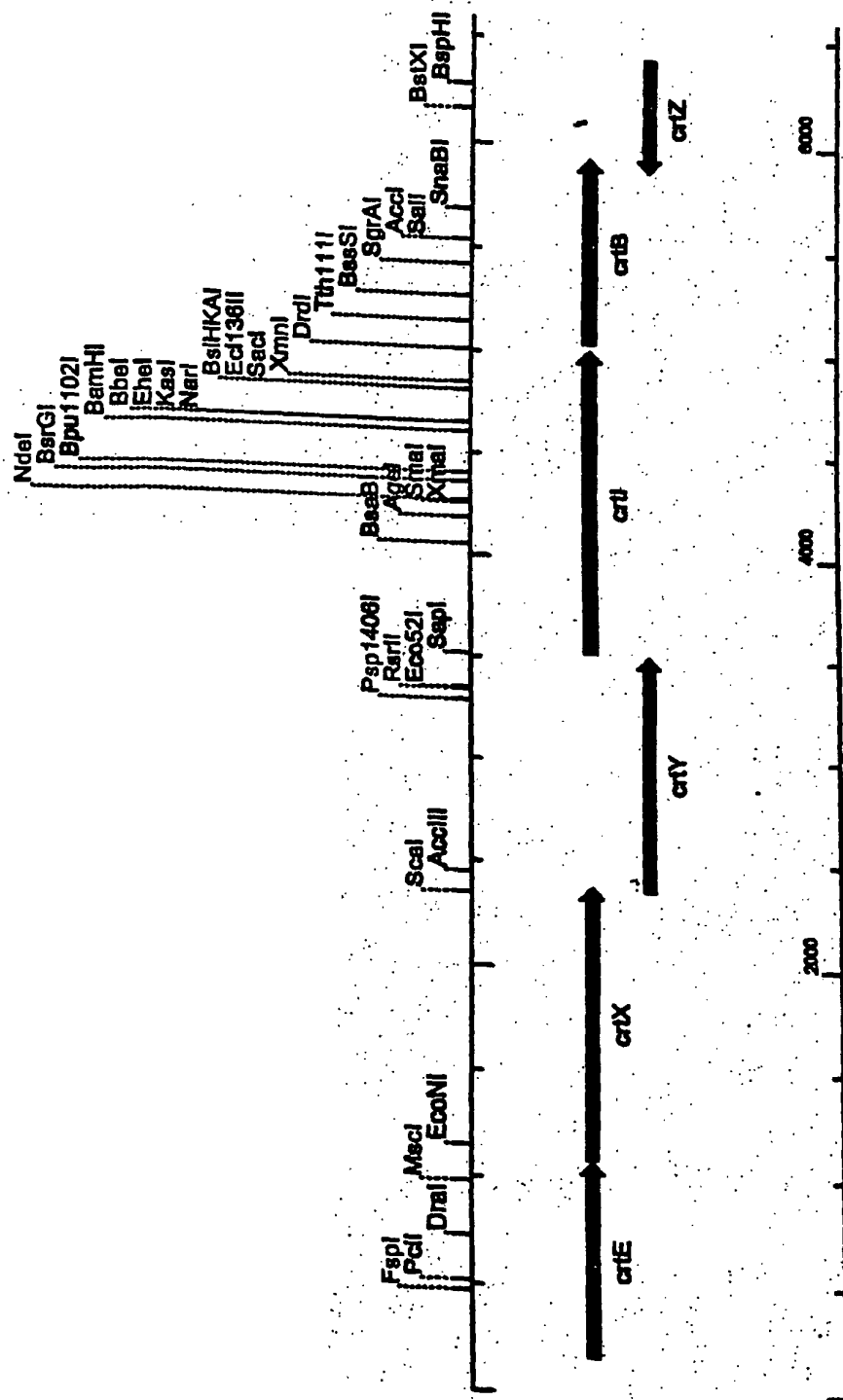


Figure 2



*P. stewartii* Operon (6586 bps)



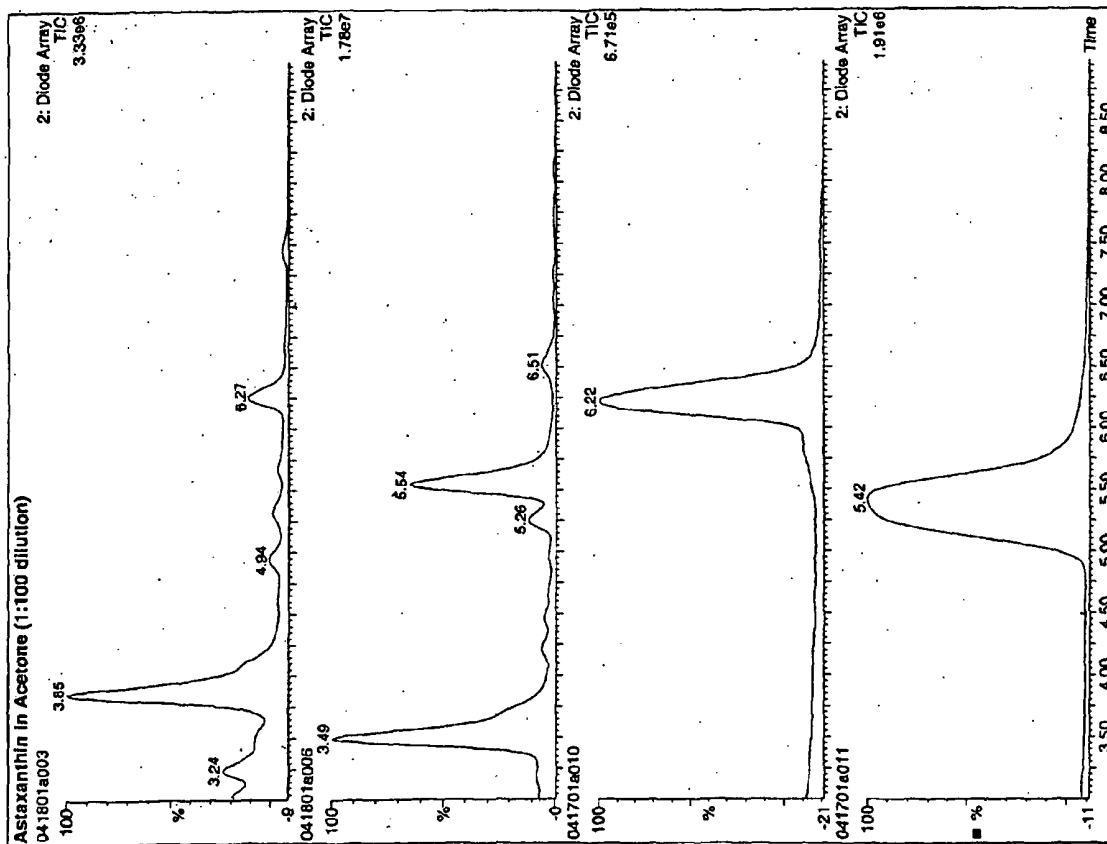
# HPLC Analysis

*Pantoea stewartii* – zeaxanthin production

*P. stewartii*: *crtW* (*Brevundimonas aurantiaca*)  
– astaxanthin production

Zeaxanthin standard

Astaxanthin standard



## SEQUENCE LISTING

<110> Cargill, Incorporated

<120> Carotenoid Biosynthesis

<130> 12794-004W01

<150> US 60/288,984

<151> 2001-05-04

<150> US 60/264,329

<151> 2001-01-26

<160> 47

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1296

<212> DNA

<213> *Pantoea stewartii*

<400> 1

atgagccatt	ttgcggtgat	cgcaccgccc	tttttcagcc	atgttcgcgc	tctgcaaaac	60
cttgctcagg	aattagtggc	ccgcggtcat	cgtgttacgt	tttttcagca	acatgactgc	120
aaagcgctgg	taacgggcag	cgatatcgga	ttccagaccg	tcggactgca	aacgcatect	180
cccggttcct	tatcgcacct	gctgcacctg	gccgcgcacc	cactcggacc	ctcgatgtta	240
cgactgatca	atgaaatggc	acgtaccagc	gatatgcttt	gccgggaact	gcccgccgct	300
tttcatgcgt	tgcagataga	gggcgtgatc	gttgatcaaa	tggagccggc	aggtgcagta	360
gtcgcagaag	cgtcagggtc	gccgtttgtt	tcggtggcct	gcgcgctgcc	gctcaaccgc	420
gaaccggggt	tgctcttggc	ggtgatgcct	ttcgagtacg	gcaccagcga	tgccggctcg	480
gaacgctata	ccaccagcga	aaaaatttat	gactggctga	tgcgacgtca	cgatcgtgtg	540
atcgcgcatc	atgcatgcag	aatgggttta	gccccgcgtg	aaaaactgca	tcattgtttt	600
tctccactgg	cacaaatcag	ccagttgatc	cccgaactgg	attttccccc	caaagcgctg	660
ccagactgct	ttcatgcggt	tggaccgtta	cggcaacccc	aggggacgcc	ggggtcacat	720
acttcttatt	ttccgtcccc	ggacaaaccc	cgtatttttg	cctcgctggg	caccctgcag	780
ggacatcggt	atggcctggt	caggaccatc	gccaaagcct	gcgaagaggt	ggatgcgcag	840
ttactgttgg	cacactgtgg	cggcctctca	gccacgcagg	caggtgaact	ggcccggggc	900
ggggacattc	aggttggtga	ttttgccgat	caatccgcag	cactttcaca	ggcacagttg	960
acaatcacac	atggtgggat	gaatacggta	ctggacgcta	ttgcttccc	cacaccgcta	1020
ctggcgctgc	cgctggcatt	tgatcaacct	ggcgtggcat	cacgaattgt	ttatcatggc	1080
atcggcaagc	gtgcgtctcg	gtttactacc	agccatgcgc	tggcgcggca	gattcgatcg	1140
ctgctgacta	acaccgatta	cccgcagcgt	atgacaaaaa	ttcaggccgc	attgcgtctg	1200
gcaggcggca	caccagccgc	cgccgatatt	gttgaacagg	cgatgcggac	ctgtcagcca	1260
gtactcagtg	ggcaggatta	tgcaaccgca	ctatga			1296

<210> 2

<211> 431

<212> PRT

<213> *Pantoea stewartii*

<400> 2

Met Ser His Phe Ala Val Ile Ala Pro Pro Phe Phe Ser His Val Arg

1

5

10

15

```

Ala Leu Gln Asn Leu Ala Gln Glu Leu Val Ala Arg Gly His Arg Val
      20      25      30
Thr Phe Phe Gln Gln His Asp Cys Lys Ala Leu Val Thr Gly Ser Asp
      35      40      45
Ile Gly Phe Gln Thr Val Gly Leu Gln Thr His Pro Pro Gly Ser Leu
      50      55      60
Ser His Leu Leu His Leu Ala Ala His Pro Leu Gly Pro Ser Met Leu
      65      70      75      80
Arg Leu Ile Asn Glu Met Ala Arg Thr Ser Asp Met Leu Cys Arg Glu
      85      90      95
Leu Pro Ala Ala Phe His Ala Leu Gln Ile Glu Gly Val Ile Val Asp
      100      105      110
Gln Met Glu Pro Ala Gly Ala Val Ala Glu Ala Ser Gly Leu Pro
      115      120      125
Phe Val Ser Val Ala Cys Ala Leu Pro Leu Asn Arg Glu Pro Gly Leu
      130      135      140
Pro Leu Ala Val Met Pro Phe Glu Tyr Gly Thr Ser Asp Ala Ala Arg
      145      150      155      160
Glu Arg Tyr Thr Thr Ser Glu Lys Ile Tyr Asp Trp Leu Met Arg Arg
      165      170      175
His Asp Arg Val Ile Ala His His Ala Cys Arg Met Gly Leu Ala Pro
      180      185      190
Arg Glu Lys Leu His His Cys Phe Ser Pro Leu Ala Gln Ile Ser Gln
      195      200      205
Leu Ile Pro Glu Leu Asp Phe Pro Arg Lys Ala Leu Pro Asp Cys Phe
      210      215      220
His Ala Val Gly Pro Leu Arg Gln Pro Gln Gly Thr Pro Gly Ser Ser
      225      230      235      240
Thr Ser Tyr Phe Pro Ser Pro Asp Lys Pro Arg Ile Phe Ala Ser Leu
      245      250      255
Gly Thr Leu Gln Gly His Arg Tyr Gly Leu Phe Arg Thr Ile Ala Lys
      260      265      270
Ala Cys Glu Glu Val Asp Ala Gln Leu Leu Leu Ala His Cys Gly Gly
      275      280      285
Leu Ser Ala Thr Gln Ala Gly Glu Leu Ala Arg Gly Gly Asp Ile Gln
      290      295      300
Val Val Asp Phe Ala Asp Gln Ser Ala Ala Leu Ser Gln Ala Gln Leu
      305      310      315      320
Thr Ile Thr His Gly Gly Met Asn Thr Val Leu Asp Ala Ile Ala Ser
      325      330      335
Arg Thr Pro Leu Leu Ala Leu Pro Leu Ala Phe Asp Gln Pro Gly Val
      340      345      350
Ala Ser Arg Ile Val Tyr His Gly Ile Gly Lys Arg Ala Ser Arg Phe
      355      360      365
Thr Thr Ser His Ala Leu Ala Arg Gln Ile Arg Ser Leu Leu Thr Asn
      370      375      380
Thr Asp Tyr Pro Gln Arg Met Thr Lys Ile Gln Ala Ala Leu Arg Leu
      385      390      395      400
Ala Gly Gly Thr Pro Ala Ala Ala Asp Ile Val Glu Gln Ala Met Arg
      405      410      415
Thr Cys Gln Pro Val Leu Ser Gly Gln Asp Tyr Ala Thr Ala Leu
      420      425      430

```

&lt;210&gt; 3

&lt;211&gt; 1149

&lt;212&gt; DNA

&lt;213&gt; Pantoea stewartii

&lt;400&gt; 3

```

atgcaaccgc actatgatct cattctggtc ggtgccggtc tggctaattgg ccttatcgcg      60
ctccggcttc agcaacagca tccggatatg cggatcttgc ttattgaggc gggtcctgag      120
gcgggaggga accatactcg gtcctttcac gaagaggatt taacgctgaa tcagcatcgc      180
tggtatagcgc cgcttgtggt ccatcactgg cccgactacc aggttcgttt cccccaacgc      240
cgtcgccatg tgaacagtgg ctactactgc gtgacctccc ggcatttcgc cgggatactc      300
cggcaacagt ttggacaaca tttatggctg cataccgcgg tttcagccgt tcatgctgaa      360
tcggtccagt tagcggatgg ccggattatt catgccagta cagtgatcga cggacggggt      420
tacacgcctg attctgcact acgcgtagga ttccaggcat ttatcggtea ggagtggcaa      480
ctgagcgcgc cgcattggtt atcgtcaccg attatcatgg atgcgacggt cgatcagcaa      540
aatggctacc gctttgttta taccctgccg ctttccgcaa ccgcactgct gatcgaagac      600
acacactaca ttgacaaggc taatcttcag gccgaacggg cgcgtcagaa cattcgcgat      660
tatgctgcgc gacagggttg gccgttacag acgttgctgc gggaagaaca ggggtgcattg      720
cccattacgt taacgggcga taatcgtcag ttttggaac agcaaccgca agcctgtagc      780
ggattacgcg ccgggctgtt tcatccgaca accggctact ccctaccgct cgcggtggcg      840
ctggccgatc gtctcagcgc gctggatgtg tttacctctt cctctgttca ccagacgatt      900
gctcactttg cccagcaacg ttggcagcaa caggggtttt tccgcattgct gaatcgcattg      960
ttgttttttag ccggaccggc cgagtcacgc tggcgtgtga tgcagcggtt ctatggctta      1020
cccaggagatt tgattgcccg cttttatgcg ggaaaactca ccgtgaccga tcggctacgc      1080
attctgagcg gcaagccgcc cgttcccgtt ttcgcggcat tgcaggcaat tatgacgact      1140
catcggtga

```

&lt;210&gt; 4

&lt;211&gt; 382

&lt;212&gt; PRT

&lt;213&gt; Pantoea stewartii

&lt;400&gt; 4

```

Met Gln Pro His Tyr Asp Leu Ile Leu Val Gly Ala Gly Leu Ala Asn
 1          5          10          15
Gly Leu Ile Ala Leu Arg Leu Gln Gln Gln His Pro Asp Met Arg Ile
 20          25          30
Leu Leu Ile Glu Ala Gly Pro Glu Ala Gly Gly Asn His Thr Trp Ser
 35          40          45
Phe His Glu Glu Asp Leu Thr Leu Asn Gln His Arg Trp Ile Ala Pro
 50          55          60
Leu Val Val His His Trp Pro Asp Tyr Gln Val Arg Phe Pro Gln Arg
 65          70          75          80
Arg Arg His Val Asn Ser Gly Tyr Tyr Cys Val Thr Ser Arg His Phe
 85          90          95
Ala Gly Ile Leu Arg Gln Gln Phe Gly Gln His Leu Trp Leu His Thr
 100         105         110
Ala Val Ser Ala Val His Ala Glu Ser Val Gln Leu Ala Asp Gly Arg
 115         120         125
Ile Ile His Ala Ser Thr Val Ile Asp Gly Arg Gly Tyr Thr Pro Asp
 130         135         140
Ser Ala Leu Arg Val Gly Phe Gln Ala Phe Ile Gly Gln Glu Trp Gln
 145         150         155         160
Leu Ser Ala Pro His Gly Leu Ser Ser Pro Ile Ile Met Asp Ala Thr
 165         170         175
Val Asp Gln Gln Asn Gly Tyr Arg Phe Val Tyr Thr Leu Pro Leu Ser
 180         185         190
Ala Thr Ala Leu Leu Ile Glu Asp Thr His Tyr Ile Asp Lys Ala Asn
 195         200         205
Leu Gln Ala Glu Arg Ala Arg Gln Asn Ile Arg Asp Tyr Ala Ala Arg
 210         215         220

```

Gln Gly Trp Pro Leu Gln Thr Leu Leu Arg Glu Glu Gln Gly Ala Leu  
 225 230 235 240  
 Pro Ile Thr Leu Thr Gly Asp Asn Arg Gln Phe Trp Gln Gln Gln Pro  
 245 250 255  
 Gln Ala Cys Ser Gly Leu Arg Ala Gly Leu Phe His Pro Thr Thr Gly  
 260 265 270  
 Tyr Ser Leu Pro Leu Ala Val Ala Leu Ala Asp Arg Leu Ser Ala Leu  
 275 280 285  
 Asp Val Phe Thr Ser Ser Ser Val His Gln Thr Ile Ala His Phe Ala  
 290 295 300  
 Gln Gln Arg Trp Gln Gln Gln Gly Phe Phe Arg Met Leu Asn Arg Met  
 305 310 315 320  
 Leu Phe Leu Ala Gly Pro Ala Glu Ser Arg Trp Arg Val Met Gln Arg  
 325 330 335  
 Phe Tyr Gly Leu Pro Glu Asp Leu Ile Ala Arg Phe Tyr Ala Gly Lys  
 340 345 350  
 Leu Thr Val Thr Asp Arg Leu Arg Ile Leu Ser Gly Lys Pro Pro Val  
 355 360 365  
 Pro Val Phe Ala Ala Leu Gln Ala Ile Met Thr Thr His Arg  
 370 375 380

<210> 5  
 <211> 912  
 <212> DNA  
 <213> *Pantoea stewartii*

<400> 5  
 atgatggtct ggcgcaaaaa acacgttcac cttactggca tttcggctga gcagttgctg 60  
 gctgatatcg atagccgcct tgatcagtta ctgccggttc aggggtgagcg ggattgtgtg 120  
 ggtgccgcga tgcgtgaagg cacgctggca ccgggcaaac gtattcgtcc gatgctgctg 180  
 ttattaacag cgcgcgatct tggctgtgcg atcagtcacg ggggattact ggatttagcc 240  
 tgcgcggttg aaatggtgca tgctgcctcg ctgattctgg atgatatgcc ctgcatggac 300  
 gatgcgcaga tgcgtcgggg gcgtcccacc attcacacgc agtacggtga acatgtggcg 360  
 attctggcgg cggtcgcttt actcagcaaa gcgtttgggg tgattgccga ggctgaaggt 420  
 ctgacgccga tagccaaaac tcgcgcggtg tcggagctgt ccaactgcgat tggcatgcag 480  
 ggtctggttc agggccagtt taaggacctc tcggaaggcg ataaaccccg cagcgccgat 540  
 gccatactgc taaccaatca gtttaaaacc agcacgctgt tttgcgcgtc aacgcaaattg 600  
 gcgtccattg cggccaacgc gtctgcgaa gcgcgtgaga acctgcatcg tttctcgctc 660  
 gatctcggcc aggcctttca gttgcttgac gatcttaccg atggcatgac cgataccggc 720  
 aaagacatca atcaggatgc aggtaaatca acgctggtca atttattagg ctcaggcgcg 780  
 gtcgaagaac gcctgcgaca gcatttgcg ctggccagtg aacacctttc cgcggcatgc 840  
 caaacggcc attccaccac ccaacttttt attcaggcct ggtttgacaa aaaactcgct 900  
 gccgtcagtt aa 912

<210> 6  
 <211> 303  
 <212> PRT  
 <213> *Pantoea stewartii*

<400> 6  
 Met Met Val Cys Ala Lys Lys His Val His Leu Thr Gly Ile Ser Ala  
 1 5 10 15  
 Glu Gln Leu Leu Ala Asp Ile Asp Ser Arg Leu Asp Gln Leu Leu Pro  
 20 25 30  
 Val Gln Gly Glu Arg Asp Cys Val Gly Ala Ala Met Arg Glu Gly Thr  
 35 40 45  
 Leu Ala Pro Gly Lys Arg Ile Arg Pro Met Leu Leu Leu Leu Thr Ala

50	55	60
Arg Asp Leu Gly Cys Ala	Ile Ser His Gly Gly	Leu Leu Asp Leu Ala
65	70	75
Cys Ala Val Glu Met Val	His Ala Ala Ser Leu	Ile Leu Asp Asp Met
85	90	95
Pro Cys Met Asp Asp	Ala Gln Met Arg Arg	Gly Arg Pro Thr Ile His
100	105	110
Thr Gln Tyr Gly Glu His	Val Ala Ile Leu Ala	Ala Val Ala Leu Leu
115	120	125
Ser Lys Ala Phe Gly Val	Ile Ala Glu Ala Glu	Gly Leu Thr Pro Ile
130	135	140
Ala Lys Thr Arg Ala Val	Ser Glu Leu Ser Thr	Ala Ile Gly Met Gln
145	150	155
Gly Leu Val Gln Gly Gln	Phe Lys Asp Leu Ser	Glu Gly Asp Lys Pro
165	170	175
Arg Ser Ala Asp Ala Ile	Leu Leu Thr Asn Gln	Phe Lys Thr Ser Thr
180	185	190
Leu Phe Cys Ala Ser Thr	Gln Met Ala Ser Ile	Ala Ala Asn Ala Ser
195	200	205
Cys Glu Ala Arg Glu Asn	Leu His Arg Phe Ser	Leu Asp Leu Gly Gln
210	215	220
Ala Phe Gln Leu Leu Asp	Asp Leu Thr Asp Gly	Met Thr Asp Thr Gly
225	230	235
Lys Asp Ile Asn Gln Asp	Ala Gly Lys Ser Thr	Leu Val Asn Leu Leu
245	250	255
Gly Ser Gly Ala Val Glu	Glu Arg Leu Arg Gln	His Leu Arg Leu Ala
260	265	270
Ser Glu His Leu Ser Ala	Ala Cys Gln Asn Gly	His Ser Thr Thr Gln
275	280	285
Leu Phe Ile Gln Ala Trp	Phe Asp Lys Lys Leu	Ala Ala Val Ser
290	295	300

&lt;210&gt; 7

&lt;211&gt; 1479

&lt;212&gt; DNA

<213> *Pantoea stewartii*

&lt;400&gt; 7

atgaaaccaa	ctacggtaat	tggtgcgggc	tttgggtggcc	tggcactggc	aattcgttta	60
caggccgcag	gtattcctgt	tttgctgctt	gagcagcgcg	acaagccggg	tggccgggct	120
tatgtttatc	aggagcaggg	ctttactttt	gatgcaggcc	ctaccgttat	caccgatccc	180
agcgcgattg	aagaactggt	tgctctggcc	ggtaaacagc	ttaaggatta	cgtcgagctg	240
ttgccgggtca	cgccgtttta	tcgcctgtgc	tgggagtcgg	gcaaggtctt	caattacgat	300
aacgaccagg	cccagttaga	agcgcagata	cagcagttta	atccgcgcga	tgttgcgggg	360
tatcgagcgt	tccttgacta	ttcgcggtgc	gtattcaatg	agggctatct	gaagctcggc	420
actgtgcctt	ttttatcggt	caaagacatg	cttcggggccg	cgccccagtt	ggcaaagctg	480
caggcatggc	gcagcgttta	cagtaaagtt	gccgggtaca	ttgaggatga	gcatcttcgg	540
caggcgtttt	cttttcactc	gctcttagtg	ggggggaatc	cgtttgcaac	ctcgtccatt	600
tatacgctga	ttcacgcgtt	agaacgggaa	tggggcgctc	ggtttccacg	cggtggaacc	660
ggtgcgctgg	tcaatggcat	gatcaagctg	tttcaggatc	tggggcgcg	agtcgtgctt	720
aacgcccggg	tcagtcatat	ggaaaccggt	ggggacaaga	ttcaggccgt	gcagttggaa	780
gacggcagac	ggtttgaaac	ctgcgcggtg	gcgtcgaaac	ctgatgttgt	acatacctat	840
cgcgatctgc	tgtctcagca	tcccgcagcc	gctaagcagg	cgaaaaaact	gcaatccaag	900
cgtagtagta	actcactggt	tgtactctat	tttgggtctc	accatcatca	cgatcaactc	960
gcccatacata	ccgtctgttt	tggggccacgc	taccgtgaac	tgattcacga	aatttttaac	1020
catgatggtc	tggctgagga	tttttcgctt	tatttacacg	cacctgtgtg	cacggatccg	1080
tcactggcac	cggaagggtg	cggcagctat	tatgtgctgg	cgccgtgttc	acacttaggc	1140

```

acggcgaacc tcgactgggc ggtagaagga ccccgactgc gcgatcgat ttttgactac 1200
cttgagcaac attacatgcc tggcttgcca agccagttgg tgacgcaccg tatgtttacg 1260
ccgttcgatt tccgcgacga gctcaatgcc tggcaagggt cggccttctc ggttgaacct 1320
attctgaccc agagcgcctg gttccgacca cataaccgcg ataagcacat tgataatctt 1380
tatctgggtg gcgcaggcac ccatacctggc gcgggcattc ccggcgtaat cggctcggcg 1440
aaggcgacgg caggcttaat gctggaggac ctgatttga 1479

```

&lt;210&gt; 8

&lt;211&gt; 492

&lt;212&gt; PRT

<213> *Pantoea stewartii*

&lt;400&gt; 8

```

Met Lys Pro Thr Thr Val Ile Gly Ala Gly Phe Gly Gly Leu Ala Leu
1      5      10      15
Ala Ile Arg Leu Gln Ala Ala Gly Ile Pro Val Leu Leu Leu Gln
20     25     30
Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Tyr Gln Glu Gln Gly Phe
35     40     45
Thr Phe Asp Ala Gly Pro Thr Val Ile Thr Asp Pro Ser Ala Ile Glu
50     55     60
Glu Leu Phe Ala Leu Ala Gly Lys Gln Leu Lys Asp Tyr Val Glu Leu
65     70     75     80
Leu Pro Val Thr Pro Phe Tyr Arg Leu Cys Trp Glu Ser Gly Lys Val
85     90     95
Phe Asn Tyr Asp Asn Asp Gln Ala Gln Leu Glu Ala Gln Ile Gln Gln
100    105    110
Phe Asn Pro Arg Asp Val Ala Gly Tyr Arg Ala Phe Leu Asp Tyr Ser
115    120    125
Arg Ala Val Phe Asn Glu Gly Tyr Leu Lys Leu Gly Thr Val Pro Phe
130    135    140
Leu Ser Phe Lys Asp Met Leu Arg Ala Ala Pro Gln Leu Ala Lys Leu
145    150    155    160
Gln Ala Trp Arg Ser Val Tyr Ser Lys Val Ala Gly Tyr Ile Glu Asp
165    170    175
Glu His Leu Arg Gln Ala Phe Ser Phe His Ser Leu Leu Val Gly Gly
180    185    190
Asn Pro Phe Ala Thr Ser Ser Ile Tyr Thr Leu Ile His Ala Leu Glu
195    200    205
Arg Glu Trp Gly Val Trp Phe Pro Arg Gly Gly Thr Gly Ala Leu Val
210    215    220
Asn Gly Met Ile Lys Leu Phe Gln Asp Leu Gly Gly Glu Val Val Leu
225    230    235    240
Asn Ala Arg Val Ser His Met Glu Thr Val Gly Asp Lys Ile Gln Ala
245    250    255
Val Gln Leu Glu Asp Gly Arg Arg Phe Glu Thr Cys Ala Val Ala Ser
260    265    270
Asn Ala Asp Val Val His Thr Tyr Arg Asp Leu Leu Ser Gln His Pro
275    280    285
Ala Ala Ala Lys Gln Ala Lys Lys Leu Gln Ser Lys Arg Met Ser Asn
290    295    300
Ser Leu Phe Val Leu Tyr Phe Gly Leu Asn His His His Asp Gln Leu
305    310    315    320
Ala His His Thr Val Cys Phe Gly Pro Arg Tyr Arg Glu Leu Ile His
325    330    335
Glu Ile Phe Asn His Asp Gly Leu Ala Glu Asp Phe Ser Leu Tyr Leu
340    345    350

```

His Ala Pro Cys Val Thr Asp Pro Ser Leu Ala Pro Glu Gly Cys Gly  
 355 360 365  
 Ser Tyr Tyr Val Leu Ala Pro Val Pro His Leu Gly Thr Ala Asn Leu  
 370 375 380  
 Asp Trp Ala Val Glu Gly Pro Arg Leu Arg Asp Arg Ile Phe Asp Tyr  
 385 390 395 400  
 Leu Glu Gln His Tyr Met Pro Gly Leu Arg Ser Gln Leu Val Thr His  
 405 410 415  
 Arg Met Phe Thr Pro Phe Asp Phe Arg Asp Glu Leu Asn Ala Trp Gln  
 420 425 430  
 Gly Ser Ala Phe Ser Val Glu Pro Ile Leu Thr Gln Ser Ala Trp Phe  
 435 440 445  
 Arg Pro His Asn Arg Asp Lys His Ile Asp Asn Leu Tyr Leu Val Gly  
 450 455 460  
 Ala Gly Thr His Pro Gly Ala Gly Ile Pro Gly Val Ile Gly Ser Ala  
 465 470 475 480  
 Lys Ala Thr Ala Gly Leu Met Leu Glu Asp Leu Ile  
 485 490

&lt;210&gt; 9

&lt;211&gt; 893

&lt;212&gt; DNA

<213> *Pantoea stewartii*

&lt;400&gt; 9

ccatggcgggt	tggctcgaaa	agctttgcca	ctgcatcgac	gcttttcgac	gccaaaaccc	60
gtcgcagcgt	gctgatgctt	tacgcatggt	gccgccactg	cgacgacgtc	attgacgatc	120
aaacactggg	ctttcatgcc	gaccagccct	cttcgcagat	gcctgagcag	cgccctgcagc	180
agcttgaaat	gaaaacgcgt	caggcctacg	ccggttcgca	aatgcacgag	cccgttttg	240
ccgcgtttca	ggaggtcgcg	atggcgcatg	atctcgctcc	cgccctacgcg	ttcgaccatc	300
tgggaaggttt	tgccatggat	gtgcgcgaaa	cgcgctacct	gacactggac	gatacgctgc	360
gttattgcta	tcacgtcgcc	ggtgttggtg	gcctgatgat	ggcgcaaatt	atgggcgttc	420
gcgataacgc	cacgctcgat	cgcgccctgcg	atctcgggct	ggctttccag	ttgaccaaca	480
ttgcgcgtga	tattgtcgac	gatgctcagg	tgggccgctg	ttatctgcct	gaaagctggc	540
tgggaagagga	aggactgacg	aaagcgaatt	atgctgcgcc	agaaaaccgg	caggccttaa	600
gccgtatcgc	cgggcgactg	gtacgggaag	gggaacccta	ttacgtatca	tcaatggccg	660
gtctggcaca	attaccctta	cgctcggcct	gggccatcgc	gacagcgaag	cagggtgtacc	720
gtaaaatttg	cgtgaaagtt	gaacaggccg	gtaagcaggc	ctgggatcat	cgccagtcca	780
cgtccaccgc	cgaaaaatta	acgcttttgc	tgacggcatc	cggtcaggca	gttacttccc	840
ggatgaagac	gtatccaccc	cgtcctgctc	atctctggca	gcgcccgatc	tag	893

&lt;210&gt; 10

&lt;211&gt; 296

&lt;212&gt; PRT

<213> *Pantoea stewartii*

&lt;400&gt; 10

Met Ala Val Gly Ser Lys Ser Phe Ala Thr Ala Ser Thr Leu Phe Asp	
1 5 10 15	
Ala Lys Thr Arg Arg Ser Val Leu Met Leu Tyr Ala Trp Cys Arg His	
20 25 30	
Cys Asp Asp Val Ile Asp Asp Gln Thr Leu Gly Phe His Ala Asp Gln	
35 40 45	
Pro Ser Ser Gln Met Pro Glu Gln Arg Leu Gln Gln Leu Glu Met Lys	
50 55 60	
Thr Arg Gln Ala Tyr Ala Gly Ser Gln Met His Glu Pro Ala Phe Ala	
65 70 75 80	



Ala Phe Gln Glu Val Ala Met Ala His Asp Ile Ala Pro Ala Tyr Ala  
85 90 95  
Phe Asp His Leu Glu Gly Phe Ala Met Asp Val Arg Glu Thr Arg Tyr  
100 105 110  
Leu Thr Leu Asp Asp Thr Leu Arg Tyr Cys Tyr His Val Ala Gly Val  
115 120 125  
Val Gly Leu Met Met Ala Gln Ile Met Gly Val Arg Asp Asn Ala Thr  
130 135 140  
Leu Asp Arg Ala Cys Asp Leu Gly Leu Ala Phe Gln Leu Thr Asn Ile  
145 150 155 160  
Ala Arg Asp Ile Val Asp Asp Ala Gln Val Gly Arg Cys Tyr Leu Pro  
165 170 175  
Glu Ser Trp Leu Glu Glu Glu Gly Leu Thr Lys Ala Asn Tyr Ala Ala  
180 185 190  
Pro Glu Asn Arg Gln Ala Leu Ser Arg Ile Ala Gly Arg Leu Val Arg  
195 200 205  
Glu Ala Glu Pro Tyr Tyr Val Ser Ser Met Ala Gly Leu Ala Gln Leu  
210 215 220  
Pro Leu Arg Ser Ala Trp Ala Ile Ala Thr Ala Lys Gln Val Tyr Arg  
225 230 235 240  
Lys Ile Gly Val Lys Val Glu Gln Ala Gly Lys Gln Ala Trp Asp His  
245 250 255  
Arg Gln Ser Thr Ser Thr Ala Glu Lys Leu Thr Leu Leu Thr Ala  
260 265 270  
Ser Gly Gln Ala Val Thr Ser Arg Met Lys Thr Tyr Pro Pro Arg Pro  
275 280 285  
Ala His Leu Trp Gln Arg Pro Ile  
290 295

<210> 11  
<211> 528  
<212> DNA  
<213> *Pantoea stewartii*

<400> 11  
atgttggtgga tttggaatgc cctgatcgtg tttgtcaccg tggtcggcat ggaagtgggt 60  
gctgcactgg cacataaata catcatgcac ggctgggggt ggggctggca tctttcacat 120  
catgaaccgc gtaaaggcgc atttgaagtt aacgatctct atgccgtggt attcgccatt 180  
gtgtcgattg ccttgattta cttcggcagt acaggaatct ggccgctcca gtggattgggt 240  
gcaggcatga ccgcttatgg tttactgtat tttatggtcc acgacggact ggtacaccag 300  
cgctggccgt tccgctacat accgcgcaaa ggctacctga aacggttata catggcccac 360  
cgtatgcac atgctgtaag gggaaaagag ggctgcgtgt cctttgggtt tctgtacgcg 420  
ccaccgttat ctaaacttca ggcgacgctg agagaaaggc atgcggctag atcgggcgct 480  
gccagagatg agcaggacgg ggtggatacg tcttcacccg ggaagtaa 528

<210> 12  
<211> 175  
<212> PRT  
<213> *Pantoea stewartii*

<400> 12  
Met Leu Trp Ile Trp Asn Ala Leu Ile Val Phe Val Thr Val Val Gly  
1 5 10 15  
Met Glu Val Val Ala Ala Leu Ala His Lys Tyr Ile Met His Gly Trp  
20 25 30  
Gly Trp Gly Trp His Leu Ser His His Glu Pro Arg Lys Gly Ala Phe  
35 40 45

Glu Val Asn Asp Leu Tyr Ala Val Val Phe Ala Ile Val Ser Ile Ala  
     50                    55                    60  
 Leu Ile Tyr Phe Gly Ser Thr Gly Ile Trp Pro Leu Gln Trp Ile Gly  
     65                    70                    75                    80  
 Ala Gly Met Thr Ala Tyr Gly Leu Leu Tyr Phe Met Val His Asp Gly  
                     85                    90                    95  
 Leu Val His Gln Arg Trp Pro Phe Arg Tyr Ile Pro Arg Lys Gly Tyr  
                     100                    105                    110  
 Leu Lys Arg Leu Tyr Met Ala His Arg Met His His Ala Val Arg Gly  
                     115                    120                    125  
 Lys Glu Gly Cys Val Ser Phe Gly Phe Leu Tyr Ala Pro Pro Leu Ser  
     130                    135                    140  
 Lys Leu Gln Ala Thr Leu Arg Glu Arg His Ala Ala Arg Ser Gly Ala  
     145                    150                    155                    160  
 Ala Arg Asp Glu Gln Asp Gly Val Asp Thr Ser Ser Ser Gly Lys  
                     165                    170                    175

<210> 13  
 <211> 29  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 13  
 atyatgcacg gctggggwtg gsgmtggca

29

<210> 14  
 <211> 31  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 14  
 ggccarcgyt gatgcaccag mccgtcrtgc a

31

<210> 15  
 <211> 26  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 15  
 ctgatgctct aygcctgggtg ccgcca

26

<210> 16  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 16  
tcgcgrgcra trttsgtcar ctg 23

<210> 17  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 17  
atbmtsattg aygcsacsgt 20

<210> 18  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 18  
ytracgcarg ayacgcrcra 20

<210> 19  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 19  
rsggcagyga atagccrgtg 20

<210> 20  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 20  
aacagcatcsc grttcagcag gcgsa 25

<210> 21  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 21

ccgacggтка tcaccgatcc 20

<210> 22  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 22  
ctgcgccsac caggtagag 19

<210> 23  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 23  
ctygacgaya tgccctgcat ggac 24

<210> 24  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 24  
gtcgatttwc csgcgtcctk attg 24

<210> 25  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 25  
ggccgaattc caacgatgct ctggcagtta 30

<210> 26  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 26  
ggccagatct acttcaggcg acgctgagag 30

<210> 27  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 27  
ggccagatct tacgcgcggg taaagccaat 30

<210> 28  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 28  
ggcctctaga attaccgcgt gggtctgaag 30

<210> 29  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 29  
ggcctctaga tctgtacgcg ccaccgttat 30

<210> 30  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 30  
catcggtaag atcgtcaagc aactgaa 27

<210> 31  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 31  
gatttacctg catcctgatt gatgtct 27

<210> 32  
<211> 27

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 32  
atgtataacc gtttcaggta gcctttg 27

<210> 33  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 33  
aatacagtaa accataagcg gtcatgc 27

<210> 34  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 34  
ttcatcatcg cgcacgac 18

<210> 35  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 35  
agrtgrtgyt cgtgrtga 18

<210> 36  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 36  
gcggcatagg ctagattgaa g 21

<210> 37  
<211> 20  
<212> DNA  
<213> Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 37

gcgagttcct tctcacctat

20

&lt;210&gt; 38

&lt;211&gt; 735

&lt;212&gt; DNA

&lt;213&gt; Brevundimonas aurantiaca

&lt;400&gt; 38

atgaccgccg	ccgtcgccga	gccacgcacc	gtcccgcgcc	agacctggat	cggtctgacc	60
ctggcgggaa	tgatcgtggc	gggatgggcg	gttctgcatg	tctacggcgt	ctattttcac	120
cgatgggggc	cgttgacct	ggtgatcgcc	ccggcgatcg	tggcgggtcca	gacctggttg	180
tcggtcgggc	ttttcatcgt	cgcccatgac	gccatgtacg	gtccctggc	gccgggacgg	240
ccgcggctga	acgccgcagt	cgcccggtg	accctggggc	tctatgcggg	cttcgcgttc	300
gatcggtga	agacggcgca	ccacgcccac	cacgcgcgc	ccggcacggc	cgacgacccg	360
gattttcacg	ccccggcgcc	ccgcgccttc	cttccttggt	tcctgaactt	ctttcgaccc	420
tatttcggct	ggcgcgagat	ggcggctctg	accgccttgg	tcctgatcgc	cctcttcggc	480
ctggggggcg	ggccggccaa	tctcctgacc	ttctggggcg	cgccggccct	gctttcagcg	540
cttcagctct	tcaccttcgg	cacctggctg	ccgcaccgcc	acaccgacca	gccgttcgcc	600
gacgcgcacc	acgccgcag	cagcggtac	ggccccgtgc	tttccttgct	cacctgtttc	660
cacttcggcc	gccaccacga	acaccatctg	agccccctggc	ggccctggtg	gcgtctgtgg	720
cgcgcgaggt	cttga					735

&lt;210&gt; 39

&lt;211&gt; 244

&lt;212&gt; PRT

&lt;213&gt; Brevundimonas aurantiaca

&lt;400&gt; 39

Met	Thr	Ala	Ala	Val	Ala	Glu	Pro	Arg	Thr	Val	Pro	Arg	Gln	Thr	Trp
1				5					10					15	
Ile	Gly	Leu	Thr	Leu	Ala	Gly	Met	Ile	Val	Ala	Gly	Trp	Ala	Val	Leu
		20					25						30		
His	Val	Tyr	Gly	Val	Tyr	Phe	His	Arg	Trp	Gly	Pro	Leu	Thr	Leu	Val
		35					40					45			
Ile	Ala	Pro	Ala	Ile	Val	Ala	Val	Gln	Thr	Trp	Leu	Ser	Val	Gly	Leu
		50				55					60				
Phe	Ile	Val	Ala	His	Asp	Ala	Met	Tyr	Gly	Ser	Leu	Ala	Pro	Gly	Arg
65				70					75					80	
Pro	Arg	Leu	Asn	Ala	Ala	Val	Gly	Arg	Leu	Thr	Leu	Gly	Leu	Tyr	Ala
			85					90						95	
Gly	Phe	Arg	Phe	Asp	Arg	Leu	Lys	Thr	Ala	His	His	Ala	His	His	Ala
			100				105						110		
Ala	Pro	Gly	Thr	Ala	Asp	Asp	Pro	Asp	Phe	His	Ala	Pro	Ala	Pro	Arg
		115					120					125			
Ala	Phe	Leu	Pro	Trp	Phe	Leu	Asn	Phe	Phe	Arg	Thr	Tyr	Phe	Gly	Trp
		130				135					140				
Arg	Glu	Met	Ala	Val	Leu	Thr	Ala	Leu	Val	Leu	Ile	Ala	Leu	Phe	Gly
145				150					155					160	
Leu	Gly	Ala	Arg	Pro	Ala	Asn	Leu	Leu	Thr	Phe	Trp	Ala	Ala	Pro	Ala
			165					170						175	
Leu	Leu	Ser	Ala	Leu	Gln	Leu	Phe	Thr	Phe	Gly	Thr	Trp	Leu	Pro	His
			180					185						190	

Arg	His	Thr	Asp	Gln	Pro	Phe	Ala	Asp	Ala	His	His	Ala	Arg	Ser	Ser
		195					200					205			
Gly	Tyr	Gly	Pro	Val	Leu	Ser	Leu	Leu	Thr	Cys	Phe	His	Phe	Gly	Arg
	210					215					220				
His	His	Glu	His	His	Leu	Ser	Pro	Trp	Arg	Pro	Trp	Trp	Arg	Leu	Trp
225					230					235					240
Arg	Gly	Glu	Ser												

<210> 40  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 40  
 ccaygaygay atwatgga 18

<210> 41  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 41  
 yttyttvccy tycctaata 18

<210> 42  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 42  
 acagcgttgg acactcag 18

<210> 43  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 43  
 gcgtcgataa tggaagtgag 20

<210> 44  
 <211> 1496  
 <212> DNA  
 <213> Sulfolobus shibatae



<400> 44  
 ttaccagtgt taaaaagtgc tatagaaggt aaggaaagtt tagaacaatt ctttagaaag 60  
 ataataattg aattgaaggc cgccatgatg cttactgggt ctaaagacgt tgatgcgta 120  
 aagaagacca gtattgttat tttaggtaaa cttaaagagt gggcagaata tagggggata 180  
 aatttatcta tatacgagaa agttagaaag agagaataaa atgagtgcg aattaagttc 240  
 gtattttaat gatatagtta acaatgtaaa ttttcatata aaaaattttg taaagagcaa 300  
 tgttagaacg cttgagggaag catcgtttca tttattttaca gctgggggca aaagacttag 360  
 acccttaatt ctggtttcat cgtcagactt aattggcggg gacaggcaaa gggcatataa 420  
 ggcagcagct gccgtggaga ttcttcacaa ctttactcta gttcatgcg atataatgga 480  
 tagggattac ctaagaagag gattaccaac tgttcatgta aagtgggggtg aaccaatggc 540  
 aatacttgca ggtgattact tacacgcaa ggcttttgaa gccttaaatg aggctctaaa 600  
 aggtcttgac ggggaatacgt tttataaggc tttttccgta tttattaatt ctattgagat 660  
 aatatcgga ggtcaagcaa tggatatgtc atttgaaaat agagtagatg taactgagga 720  
 agagtacatg caaatgataa aaggaaagac tgcgatgcta ttttcatggt ctgctgcatt 780  
 aggcgggata attaacaagg ctacgcgatg tataattaaa aatttagtcg aatatggatt 840  
 aaatctaggc atatcattcc aaatagtgga tgatatctta ggaattattg gagacaaaa 900  
 ggaattaggg aaaccagttt acagtgatat tagggaaggt aagaaaacaa ttcttggtat 960  
 aaaaacttta agtgaagcta ctgacgatga aaagaaaatt ctggtttcta cgcttgggaa 1020  
 tagggaggct aaaaaggacg atcttgagag agcgtcggaa ataataagga agtattcatt 1080  
 gcaatatgca tacaatttag ctaaaaagta ctcagatcct gcattagaac atttgcgtaa 1140  
 aattccagtt tacaatgaaa ctgctgaaaa ggctttaaaa tatctagcgc agtttaccat 1200  
 tgaaaggaga aagtaaatga gcatatcagg gatattgctt tcaattttta taccctttt 1260  
 cataagctat attacaacag tctgggtaat aagacaggca aaaaagagtg ggcttgtagg 1320  
 taaggatgta aataaaccag ataaaccgga aataccacta atgggtggga taagtataat 1380  
 agccgggttt atagcgggat ccttctcctt attactaact gatgtaagaa gtgagcgagt 1440  
 aattccatct gtaatactct cctcattgct tatagcattt cttggactat tagatg 1496

<210> 45

<211> 331

<212> PRT

<213> Sulfolobus shibatae

<400> 45

Met Ser Asp Glu Leu Ser Ser Tyr Phe Asn Asp Ile Val Asn Asn Val  
 1 5 10 15  
 Asn Phe His Ile Lys Asn Phe Val Lys Ser Asn Val Arg Thr Leu Glu  
 20 25 30  
 Glu Ala Ser Phe His Leu Phe Thr Ala Gly Gly Lys Arg Leu Arg Pro  
 35 40 45  
 Leu Ile Leu Val Ser Ser Ser Asp Leu Ile Gly Gly Asp Arg Gln Arg  
 50 55 60  
 Ala Tyr Lys Ala Ala Ala Val Glu Ile Leu His Asn Phe Thr Leu  
 65 70 75 80  
 Val His Asp Asp Ile Met Asp Arg Asp Tyr Leu Arg Arg Gly Leu Pro  
 85 90 95  
 Thr Val His Val Lys Trp Gly Glu Pro Met Ala Ile Leu Ala Gly Asp  
 100 105 110  
 Tyr Leu His Ala Lys Ala Phe Glu Ala Leu Asn Glu Ala Leu Lys Gly  
 115 120 125  
 Leu Asp Gly Asn Thr Phe Tyr Lys Ala Phe Ser Val Phe Ile Asn Ser  
 130 135 140  
 Ile Glu Ile Ile Ser Glu Gly Gln Ala Met Asp Met Ser Phe Glu Asn  
 145 150 155 160  
 Arg Val Asp Val Thr Glu Glu Glu Tyr Met Gln Met Ile Lys Gly Lys  
 165 170 175  
 Thr Ala Met Leu Phe Ser Cys Ser Ala Ala Leu Gly Gly Ile Ile Asn

				180					185					190			
Lys	Ala	Ser	Asp	Asp	Ile	Ile	Lys	Asn	Leu	Val	Glu	Tyr	Gly	Leu	Asn		
		195					200					205					
Leu	Gly	Ile	Ser	Phe	Gln	Ile	Val	Asp	Asp	Ile	Leu	Gly	Ile	Ile	Gly		
		210				215						220					
Asp	Gln	Lys	Glu	Leu	Gly	Lys	Pro	Val	Tyr	Ser	Asp	Ile	Arg	Glu	Gly		
225					230					235					240		
Lys	Lys	Thr	Ile	Leu	Val	Ile	Lys	Thr	Leu	Ser	Glu	Ala	Thr	Asp	Asp		
				245						250					255		
Glu	Lys	Lys	Ile	Leu	Val	Ser	Thr	Leu	Gly	Asn	Arg	Glu	Ala	Lys	Lys		
			260					265						270			
Asp	Asp	Leu	Glu	Arg	Ala	Ser	Glu	Ile	Ile	Arg	Lys	Tyr	Ser	Leu	Gln		
		275					280					285					
Tyr	Ala	Tyr	Asn	Leu	Ala	Lys	Lys	Tyr	Ser	Asp	Leu	Ala	Leu	Glu	His		
		290				295					300						
Leu	Arg	Lys	Ile	Pro	Val	Tyr	Asn	Glu	Thr	Ala	Glu	Lys	Ala	Leu	Lys		
305					310					315					320		
Tyr	Leu	Ala	Gln	Phe	Thr	Ile	Glu	Arg	Arg	Lys							
				325					330								

```
<210> 46
<211> 20
<212> DNA
<213> Artificial Sequence
```

<220>  
<223> Exemplary motif

<400> 46  
aggtcgtgta ctgtcagtca

20

```
<210> 47
<211> 20
<212> DNA
<213> Artificial Sequence
```

<220>  
<223> Exemplary motif

<400> 47  
acgtggtgaa ctgccagtga

20